ABSTRACT

Title of Dissertation: EPIGENETICS TUNE CHROMATIN MECHANICS, A COMPUTATIONAL APPROACH

Mary Pitman, Doctor of Philosophy, 2021

Dissertation directed by: Martin Monroe Professor Garegin A. Papoian, Department of Chemistry and Biochemistry

The base unit of DNA packaging in eukaryotes, the nucleosome, is adaptively modified for epigenetic control. Given the vast chemical space of chromatin and complexity of signaling and expression, much of our knowledge about genetic regulation comes from a biochemical or structural perspective. However, the architecture and function of chromatin also mechanically responds to non-equilibrium forces. Mechanical and biochemical properties are not independent of one another and the interplay of both of these material properties is an area of chromatin physics with many remaining questions. Therefore, I set out to determine how the material properties of chromatin are altered by biochemical variations of nucleosomes. All-atom molecular dynamics is employed coupled with new computational and theoretical tools. My findings and predictions were collaboratively validated and biologically contextualized through multiscale experimental methods.

First, I computationally discover that epigenetic switches buried within the nucleosome core alter DNA accessibility and the recruitment of essential proteins for mitosis. Next, using new computational tools, I report that centromeric nucleosomes are more elastic than their canonical counterparts and that centromeric nucleosomes rigidify when seeded for kinetochore formation. We conclude that the material properties of variants and binding events correlate with modified loading of transcriptional machinery. Further, I present my theoretical approach called Minimal Cylinder Analysis (MCA) that uses strain fluctuations to determine the Young's modulus of nucleosomes from all-atom molecular dynamics simulations. I show and explain why MCA achieves quantitative agreement with experimental measurements. Finally, the elasticity of hybrid nucleosomes in cancer is measured from simulation, and I implicate this oncogenic variant in potential neocentromere formation. Together, these data link the physics of nucleosome variations to chromatin states' plasticity and biological ramifications.

# EPIGENETICS TUNE CHROMATIN MECHANICS, A COMPUTATIONAL APPROACH

by

Mary Pitman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Garegin A. Papoian, Chair
Dr. Yamini Dalal
Professor David Fushman
Professor Pratyush Tiwary
Professor Jeffery Klauda, Dean's Representative

# Acknowledgements

It is worth reflecting on the people who made this work possible. I first thank Ellie Arroway, who inadvertently became my guide. I didn't imagine a career in science was possible. If you are a student reading this, remember that expectations are meant to be broken. A few influential scientists helped shape my career and research direction. There are numerous things I could praise my undergraduate professors for. Still, I will focus on particular moments that stand out. I thank Dr. Vadim Knyazev for his humor in class on path versus state functions ("and then you fight a pterodactyl") and his encouragement to pursue science. Dr. Gregory Miller's fresh perspective on proteins shaped my future research direction by teaching biochemistry in such a way that I walked away with questions yet unanswered. Dr. Gregory Brewer inadvertently steered me into science after complimenting a particularly embarrassing presentation. He made me realize it is okay to be awkward in science as long as your ideas are good.

During this dissertation work, Dr. Garegin Papoian's insight and energy towards science needs special thanks. I describe listening to him speak like watching a mustang—fast-paced, inspiring, and he has a way of getting people truly engaged. His methodology of seeking simple solutions first, being systematic, working smart, and life lessons were all beneficial. I also want to thank Dr. Yamini Dalal. Her devotion and passion for chromatin and biology were a significant part of the engine that powered this work. Academic debates with her were always exciting, and she

taught me a lot about the business of doing science. The impact of Garyk and Yamini's mentorship and guidance on the work in the following chapters can't be stated enough. I also thank my candidacy and thesis committee members: Dr. Dorothy Beckett, who is an exemplary woman making waves in biophysics; Dr. Pratyush Tiwary for his inspiring drive and enthusiasm; Dr. David Fushman, who is a friendly face and one of the first people I discussed chromatin fractals with; and Dr. Jeffery Klauda whose expertise in MD and technical know-how was helpful when I started doing computational modeling. I am fortunate to be surrounded by and mentored by so many gifted scientists.

While I learned a lot from advisors and senior mentors, lab mates also impacted my work and time in graduate school. I have happy memories with all of the Dalal lab members, and each person was supportive. This is reiterated in the text, but a few colleagues require particular praise. First, I thank Dr. Minh Bui, whose hands in the lab and creativity developing experimental methods are hard to beat. Also, Dr. Daniël Melters, who spent countless hours working on the 'Behemoth.' His mentorship and depth of knowledge came in clutch. Dr. Tatini Rakshit was also pivotal for her friendship and outstanding technical skills in Atomic Force Microscopy. My lab members at the University of Maryland all also had an impact on this work. I am thankful to Dr. Aravind Chandrasekaran, my batchmate and friend, for being a sounding board and comrade through this writing process and graduate school. To Dr. Cal Floyd, cheers for reviewing my code and all the good times.

I also want to highlight the contributions of friends and family. To name a few, Artur Perevalov for being one of the coolest physicists on campus, for being a trustworthy companion, and for pushing me to record and practice music together. Alex Kowler for his sarcastic and witty humor and the many times he lifted my spirits. Nishat Seraj for the rare opportunities to have 'girls' nights and her infectious confidence, and my longtime friend and musical colleague Oscar Poblete. To my family, many thanks for all the times you supported me throughout the years. Our times moving gave me the space to develop my talents and the worldview to think outside the box. Finally, I am incredibly grateful to the COVID crew of 2020-undetermined: my partner Dr. Andrew Glaudell and Dogter Miska Pitman. They both were essential in our efforts to become a three Ph.D. household. Thanks for encouraging me through this lengthy lockdown and bringing a lot of color to my time completing the writing and research in these coming chapters. Thank you to everybody!

# Table of Contents

# List of Figures

# Abbreviations

| | |
|---|---|
| AFM | Atomic Force Microscopy |
| AWL | Averaging Window Length |
| AWSEM | Associative Memory, Water Mediated, Structure and Energy Model |
| CENP-A | Centromere protein A |
| COM | Center of mass |
| FE | Free energy |
| FEA | Finite element analysis |
| FERN | Flexible Exterior Residue Number |
| H3, CENP-A, H4, H2B, H2A | Names of histone proteins |
| HAT | Histone acetyltransferase |
| HDAC | Histone deacetylase |
| IF | Immunofluorescence |
| LH | Linker histone |
| MCA | Minimal Cylinder Analysis |
| MD | Molecular dynamics |
| ML | Machine learning |
| NCP | Nucleosome core particle |
| NPT | Isothermal-isobaric ensemble |
| PDB | Protein Data Bank |
| PCA | Principal component analysis |
| PC | Principal component |
| Pre-RC | Pre-replication complex |
| PTM | Post-translational modification |
| RFC | Residue Flexibility Cutoff |
| RMSD | Root-mean-square deviation |
| RMSF | Root-mean-square fluctuation |
| RNAP2 | RNA polymerase 2 |
| TIP3P | 3-site water model |

# Chapter 1: Introduction

## 1.1 Chromatin motifs

The complex configurational landscapes of proteins arise from a palette of amino acids. Similarly, chromatin is formed from a base unit, the nucleosome, which alters eukaryotic genome organization when structurally varied. Nucleosomes, a core of histone proteins wrapped by DNA, assemble along the chromatin array and modulate genes through epigenetic modifications (1–3), histone variants (4–6), binding partners (7–9), and topological patterns (7, 10–12). From such simple tools, great biological complexity emerges. We can find an analogy to this in the arts: discrete vibrational frequencies, rhythms, and patterns generate a diverse musical catalog. Recall a song and imagine trying to transcribe the composition onto paper for long-term storage and replication. To do so, you would require a theory to categorize and contextualize the essential elements, the notes. Similarly, we must study the fundamental units of chromatin to develop a robust framework for epigenetic architecture and dynamics.

### 1.1.1 Macromolecular variations on the nucleosome

Variations in histones imbue the chromatin fiber with distinct material properties (9, 13). In eukaryotes, the proteinaceous nucleosome core contains two copies of histones from the H3, H4, H2B, and H2A families. Each histone family is

comprised of variants. The histones in the H3 family are distributed throughout the chromatin arms and form 'canonical' nucleosomes (14). In contrast, the histone H3-like variant, centromere protein A (CENP-A), is deposited primarily at the centromere (15). Linker histones (LHs), an additional histone type, bind to nucleosomes and linker DNA, compacting chromatin (8, 16, 17).

The variation in eukaryotic histones arose over long evolutionary timescales. It is thought that histone variants evolved from Archaea where instead of an octameric complex of at least four histone variants seen in eukaryotes, there is a histone scaffold formed from a repetitive monomer (18). One proposed mechanism is that divergent evolution from Archaea gave rise to H3 and H4 and later H2A and H2B (19). The eukaryotic histone variants of the H3 family—more specifically CENP-A, H3.3 and H3—are of particular interest in the coming chapters.

To compare the sequences of CENP-A, H3.3, and H3, I performed a Waterman-Eggert pairwise local alignment (20) of the structures simulated or studied in Chapters 2, 3, and 4 (Fig. 1.1a). These histone structures compared here (Fig 1.1) are the globular histone domains with external tails either cleaved or unresolved. The *Homo sapien* CENP-A histone sequence is from PDB ID: 3AN2 (15), as described in *section 2.5.1*. The *Homo sapien* H3.3 sequence is from PDB ID: 3WTP (21), as described in *section 4.2.5*. The H3 sequence compared in Fig. 1.1a is that of *Xenopus laevis*, PDB ID: 1KX5 (22), which is used as a reference in Chapter 3 from prior

simulation work by Winogradoff *et al.* (23). To differentiate between species, I will refer to the *Xenopus laevis* H3 variant as H3*.

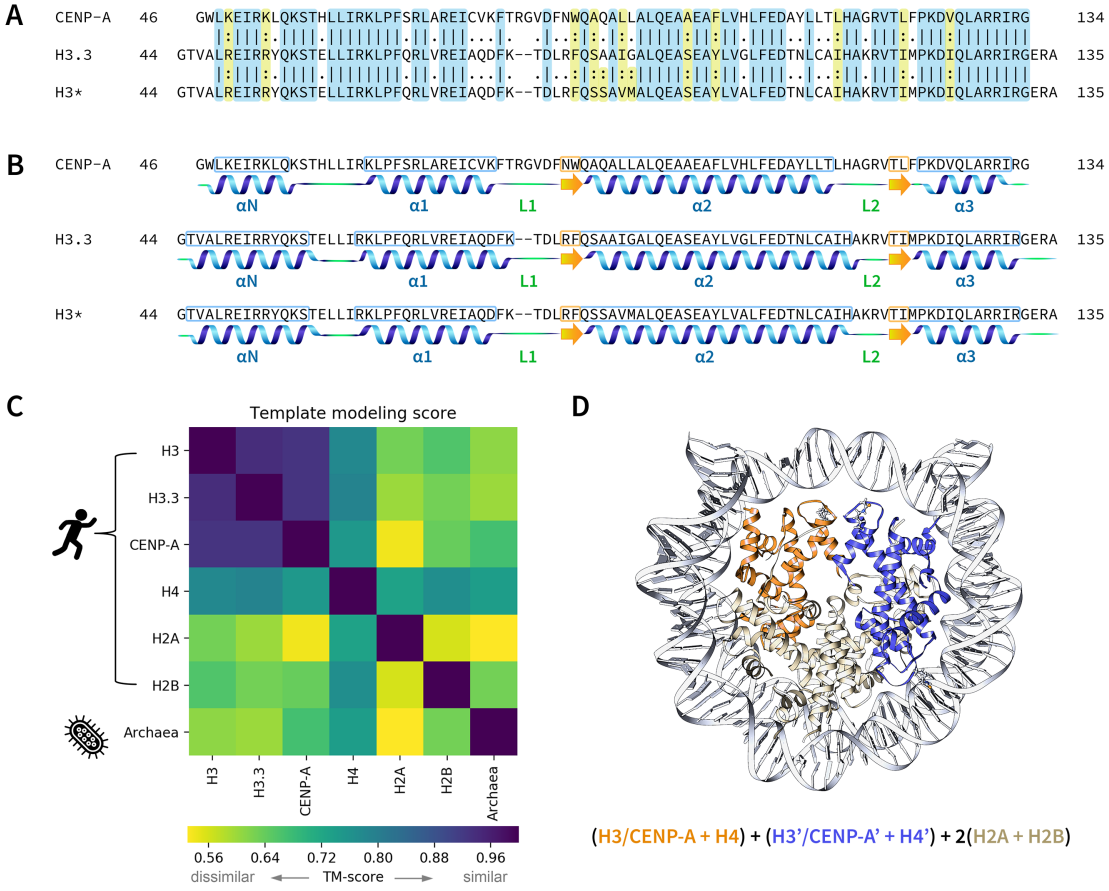


**Figure 1.1: Review of the sequences and structures of simulated histones, for reference. (a)** A sequence alignment of the globular domains of the H3 family histones studied in Chapters 2, 3, and 4. The sequences of H3.3 and H3 are compared to CENP-A. A blue highlight and a '|' symbol indicate residue identity to CENP-A. A green highlight and a ':' symbol means that the amino acids are similar to CENP-A. Residues without a highlight and with a '.' symbol are chemically dissimilar to CENP-A. **(b)** The secondary structure information of simulated H3 family histones. Depicted are α-helices as blue helices, β-sheets as orange arrows, and intrinsically disordered regions or loops as a green linear segment. Secondary structure residues are boxed. **(c)** Structural comparison of *Homo sapien* histones and the histone based chromatin of Archaea, the DNA-binding protein HMf-2 of *Methanothermus fervidus*. Crystallographically solved histone structures are compared using template modeling (TM). The histones are first TM-aligned and then TM-scored from 0 to 1, where a score of 0 indicates no chemical and structural overlap and a score of 1 is identity. **(d)** The structure of eukaryotic nucleosomes with histone tails cleaved. H3-like histones bound to H4 are shown in orange and blue. Histones H2A and H2B are shown in tan.

Standard penalties (24) for the alignment of the proteins were used (Fig. 1.1a), a penalty of 14 points for gap creation in alignment, and a penalty of 4 for each residue in the gap. A similar alignment pattern of the CENP-A globular domain compared to *Homo sapien* H3 was reported previously by Tachiwana *et al.* (15).

When comparing the sequence of CENP-A to H3.3 and H3*, many secondary structural regions are conserved (Fig. 1.1a). H3* and H3.3 have the same identity score to CENP-A (63.2%). The similarity of protein sequences, which also accounts for chemically similar mutations, can be a more meaningful metric than sequence identity. When similar and identical residues are scored, H3* is slightly more similar to CENP-A (75.9%) than H3.3 is to CENP-A (74.7%) which corresponds to one more similiarity conserved residue in H3* when compared to CENP-A. An observation is that the loop regions are points of high mutation in the simulated histone domains and could contribute to differences in observed dynamics (Fig 1.1a). The conservation of residues within the secondary structural elements of H3 histone variants agrees with the relationship between long timescale coevolution of residue contacts and the maintained fidelity of the protein fold (25). Sequence regions with higher dissimilarity to CENP-A also show differences in the lengths of α-helices found with crystallography (Fig. 1.1b).

I next calculated the structural and chemical similarity of histone variants using template modeling (TM) (Fig. 1.1c). First, the C-α atoms were TM-aligned (26) which is a method that optimizes alignment based on homology. Briefly, TM-

alignment calculates a score for each aligned residue pair based on residue and secondary structure identity with allowed alignment gaps. The initial TM-align score is then averaged with a gapless alignment of the two compared structures. Subsequently, I calculated the maximum TM-score (26):

$$\text{TM-score} = \max_{0 \le x \le 1} \left[ N_{\text{target}}^{-1} \sum_{i}^{N_{\text{aligned}}} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2} \right]$$

where $N_{Target}$ is the number of C-α atoms in the alignment target sequence, $N_{aligned}$ is the number of C-α atoms aligned, $d_i$ is the distance between the i-th pair of aligned residues, and $d_0$ normalizes the distance. TM-score is a more accurate measure of global topology than root-mean-square deviation (RMSD) which weights the distances of any residue pair equally. TM-score weights smaller distances, such as local interactions in secondary structure, more strongly than distant pairs. From this analysis, the human H3 family histones—CENP-A, H3.3, and H3 from PDB ID: 6V2K (27)—show high structural similarity, indicated by the blue box in the upper left corner (Fig. 1.1c). This analysis shows that H4 appears to have the broadest chemical and structural similarity to histone variants, including the archaeal DNA-binding protein HMf-2 of *Methanothermus fervidus* (PDB ID: 5T5K) (28).

When DNA wraps the protein octamer of H3, H4, H2A, and H2B, the nucleosome core particle is formed (Fig. 1.1d). Histone variants within the nucleosome core create local chromatin domains with distinct functions. For example, CENP-A marks the centromere for mitotic spindle attachment during mitosis. But

how is the centromere altered during mitosis? How the material properties of centromeric nucleosomes contribute to the seeding of mitosis and transcriptional plasticity is a question I tackle in the coming chapters with theoretical tools and molecular dynamics. I will further dissect how the enrichment of CENP-A at ectopic sites in cancer (29) and hybrid nucleosome containing both CENP-A and H3.3 could give rise to cancer (30). Histone or nucleosome variants also alter chromatin architecture. One example of this is that linker histones can drive chromatin compaction to form heterochromatin with diminished genetic accessibility. Fascinatingly, the inverse relationship between linker histone (LH) concentration and gene expression regulates cell stemness (8). The reduction of LH levels can ultimately lead to lymphoma (8). However, there are numerous other ways that PTMs are associated with altered chromatin architecture, such as chromatin loops formed by CTCF and condensin (31, 32) or chromatin relaxation through histone tails acetylations leading to increased increases gene expression (33). The association between PTMs and chromatin architecture has been studied with distinct classifications (31, 34).

### 1.1.2 Molecular variations on the nucleosome

Even though a cell's interior is noisy, point chemical modifications have a detectable perturbative effect due to signal amplification. The growing cascade of macromolecular interaction networks that leads to this amplification is known as signal transduction. Another mode of signal amplification is mechanotransduction—where the cell responds to stimuli due to force sensitivity. There can indeed be

feedback between chemical and mechanical events. Chemical composition calibrates the mechanical properties of biomaterials (9, 35–37), and mechanical forces can inform chemistry and *vice versa* (38, 39). An example where chemical signaling alters rigidity to regulate vital cellular processes is covered in Chapter 2. Specifically, I will interrogate how post-translational modifications (PTMs) of the centromeric histone core changes the accessibility of nucleosome bound DNA and binding motifs, aiding in centromeric replication timing (40).

The number of known histone PTM states is combinatorially expansive. In 2015 there were one hundred well studied histone PTMs cataloged and about 460 additional histone PTMs with unknown function (41). The array of known histone PTMs continues to grow with technological advancements such as improved antibody reagents and mass spectrometry methods (42). Increasingly, research is leveraging machine learning (ML) strategies to detect the effects of PTM patterns on chromatin function. One study used ML to infer promoter activity from H3K4me3 and H3K27ac data (43). Further, the Onuchic group leveraged neural networks and energy landscape theory and showed that genome fold can be predicted from epigenetic markings at a 50 kilobase resolution (34). With recent creative advances in chromatin epigenetics from a physics-based perspective, the epigenetic landscape of chromatin could be a hot area for further research. With the interplay between chemistry and mechanics of chromatin in mind, let's next explore the context in which the nucleosome resides: the nucleus and chromatin's mechanical environment.

## 1.2 Chromatin mechanics

### 1.2.1 The nuclear microenvironment

Nanoscale masses become infinitesimal in the cellular context, and nanomachines leverage elasticity, viscosity, and thermal motion for mechanical force development (44, 45). For example, chromosome movement is regulated during mitosis by frictional resistance (46–49). The cellular microenvironment has a low Reynolds number—the ratio of viscous to inertial forces is much less than unity (50). Indeed, chromatin is a component of the nucleoplasm which has a high viscosity, measured ranging from that of an oil to glycerol at 37 °C (51, 52). Thus, at low Reynolds number in the cell, force production does not generate sustained inertia and there is increased irreversible dissipation by friction.

In highly viscous media, some strategies used by active matter in the macroworld are inefficient. Purcell's scallop theorem (53) illustrates this point. Scallops use a one-hinged clapping action for propulsion through the water. In contrast, Purcell hypothesized that a one-hinge scallop with time-symmetric motion has no net displacement in a highly viscous fluid (54). The example of Purcell's scallop theorem is a thought experiment that, while informative, does not capture the whole reality of mechanics in a cell. Researchers tested Purcell's theorem by building a nano-scallop. Contrary to Purcell's expectation, the scallop swam in non-Newtonian fluids (55) such as cytoplasm (56–58). How was this so? The experimental nano-scallop is flexible and uses strain rate-dependent viscosity changes in non-Newtonian

media (55) to symmetry break its reciprocal motion and process forward. There are a few lessons from the nano-scallop to keep in mind for the coming chapters' work:

1. The cellular environment requires caution when forming biological predictions from physics and simulation.
2. Experimental validation captures nonlinear environmental factors controlled or not included in simulation and shows that well-reasoned computational results have tangible value towards understanding nature.
3. We gain helpful insight by treating chromatin as an elastic biomaterial due in part to chromatin's viscoelastic environment and in the study of the active forces it experiences.

### 1.2.2 Chromatin elasticity

Ethel Glancy D'Angelo first realized that chromatin is flexible while performing micro-manipulation experiments in 1946 (59). She stretched chromatin, saw it recoil to its original dimensions, and concluded that chromosomes have intrinsic elasticity (60, 61). Her findings established the basis for thinking about chromatin as an elastic biomaterial. The viscous nuclear microenvironment and the elastic properties of chromatin are essential for mechanotransduction (62–65). While the nucleus can adapt to external stimuli through chemical pathways, mechanical signaling mediates a rapid response. An example of the speedup gained when the cell is sensitive to mechanical forces comes from non-receptor tyrosine Src family kinases (SFK). Chemical signaling of SFKs by epidermal growth factor requires tens of seconds, while mechanotransduction by integrin-mediated force sensing requires less

than a second (62). An intrinsic mechanical sensitivity of chromatin would allow dynamic response to stimuli.

Cellular material properties regulate gene expression and phenotype. For example, a stem cell's microenvironment influences the cell fate direction (66–68), and is regulated by how the cell responds to external forces. Chowdhury *et al.* showed that the softness of embryonic stem cells dictates its microenvironment response. To examine genetic regulation by force, they used magnetic beads to apply forces to stem cells and found reduced gene expression and altered differentiation (69). Thus, issections of chromatin's elastic response could provide further insight into mechanical regulation of genetic expression.

## 1.3 Prior endeavors motivating this work

### 1.3.1 Computational efforts in the Papoian lab

There has been an ongoing effort in the Papoian lab to understand nucleosome dynamics from a high-resolution dynamical approach. One early work focused on the electrostatic properties of nucleosomes. There, Materese *et al.* found that the histone core's topology creates a more optimal neutralization of DNA charge (70). The electrostatic interplay of histones and DNA was since further proven (71). These results are essential because Materese *et al.* prove that the physics of nucleosomes is not a linear sum of DNA and protein properties. Later, Winogradoff *et al*. interrogated the structural mechanisms within CENP-A nucleosomes that differentiate

centromeric from canonical nucleosome variants (23). They found a distinct shearing motion at the interface of the CENP-A histones in comparison to H3.

The Papoian lab also studied the dynamics of histone tails in isolation. Potoyan and Papoian applied energy landscape theory to find a subtle organization within the intrinsically disordered histone tail conformational ensemble (72). Later, Winogradoff *et al*. contributed to the histone tail effort by working on the cumulative effects of tail acetylations (73). These studies all leveraged the high spatial resolution of all-atom molecular dynamics (MD) to study the effect of tails in the absence of the nucleosome and preceded the work presented here.

Through the course of my doctoral work, there has been a push towards longer timescale nucleosome simulations by using the Associative Memory, Water Mediated, Structure and Energy Model (AWSEM) (74) coarse-grained model, developed in the Papoian lab. Zhang *et al.* coupled AWSEM-MD to DNA modeling and studied DNA unwrapping from the histone core (75). Subsequently, in 2021 Wu *et al.* applied AWSEM to explain how LHs dynamically bind to nucleosomes (76)—a mechanism that was not previously well understood. However, coarse-grained modeling may miss subtle changes in dynamics due to point charge mutations and may not provide as much correspondence to experimental solvent conditions.

### 1.3.2 Experimental efforts in the Dalal lab

In one of her postdoctoral studies, Yamini Dalal unearthed that centromeric nucleosomes are unlike canonical nucleosomes, with a possible altered or unstable configuration (77). Dalal *et al.* further described how this structural variant causes centromeres' functions such as epigenetic inheritance (77). This early work incorporated biophysical and whole cell techniques and laid the foundation for future endeavors. The Dalal lab of the National Cancer Institute expanded the effort to interrogate chromosomes across interdisciplinary fields. In an innovative work on CENP-A function, Bui *et al*. found increased centromeric nucleosome plasticity through the cell-cycle *in vivo* (78). This finding was pivotal because it shed light on the structural switches necessary for centromeres to function.

Establishing that indeed biological switches cause CENP-A structural variation, Walkiewicz *et al.* applied Atomic Force Microscopy (AFM) to reveal that reconstituted CENP-A nucleosomes *in vitro* do not show large configurational changes away from the octameric nucleosome structure (79). While unmodified CENP-A nucleosomes were seen to maintain an octameric configuration, CENP-A does show more metastable configurations with a glassier free energy landscape *in silico* in comparison to H3 (9, 80). These metastable configurations may allow centromeric nucleosomes to undergo more extensive structural changes when chemically perturbed.

Of medical significance, Athwal *et al.* found that CENP-A histones are deposited in cancer away from the centromere at transcriptional hotspots (30). The chaperone network involved in this oncogenic pathway of centromeric histones was dissected by Nye *et al.* (81). The Dalal lab's prior experimental and bioinformatics works shaped the direction of the collaborations presented in the coming chapters. Through the UMD-NCI Partnership for Integrative Cancer Research, I worked as a member of this interdisciplinary effort of *in silico*, *in vitro*, and *in vivo* techniques. Selected experimental data that directly test my computational predictions and provide context for my theoretical work will be briefly discussed in Chapters 2 and 3. Extended presentations of the breadth of our collaborations are externally available (9, 40).

## 1.4 Current developments in chromatin biophysics

Since processes such as protein folding and global motions occur on the timescale of microseconds to seconds and beyond (82), mesoscale modeling is a valuable tool for chromatin computational studies. Koslover *et al.* used mesoscale chromatin modeling to show that DNA elasticity, local geometry, and linker DNA length regulate the genomic fold topology (83). Mesoscale models have also been valuable when studying the effects of binding partners. As discussed above, the nucleosome has a histone core, but it also associates with the linker histone (LH) H1 family (84). When bound, LHs induce chromatin condensation (17). One mesoscale

study revealed that LHs transform both the chromatin packing ratio and bending propensity in a LH variant-specific manner (16).

Residue level alterations inform one layer of epigenetic control. In eukaryotes, H3 nucleosomes are methylated to H3K9me3. This variant has increased nucleosome core accessibility as measured by hydrogen deuterium exchange mass spectrometry and nuclear magnetic resonance (85). The H3K9me3 variant also signals protein Swi6 binding, which then compacts and transcriptionally represses chromatin (85). Analogous to other contemporary work in the field, in Chapter 2 I will show the effects of histone core acetylation on centromeric nucleosome dynamics and signaling of binding partners (40).

Computational modeling has also clarified how chromatin architecture alters transcription. Brownian dynamics and Monte Carlo simulations were combined in 2020 and showed that transcription depends on time dependent dynamic changed to chromatin crowding instead of steady state, or equilibrated, crowding. Furthermore, the chemistry and resulting structure of genes can determine the fate of transcriptional regulation, such as upregulation or downregulation (86). In Chapter 3, we connect the elasticity of nucleosome variants to chromatin fiber compaction and correlate this effect to diminished RNA polymerase 2 (RNAP2) loading. For reference, RNAP2 transcribes DNA to RNA. These recent computational advances could allow for more direct observation *in silico* about how nucleosome variants or PTMs tune transcription. Another contemporary study stressed why we must take into account

how chromatin responds to force. Chromatin does not live in constant equilibrium during interphase but instead undergoes transcription-coupled ATP-dependent activity. Agrawal *et al.* incorporated stochastic transcriptional forces into simulations and, with this minimal feature to induce chromatin structural memory, recapitulated experimentally derived chromosome organization (87). These works further establish the feedback between mechanical forces, chromosome architecture, and gene expression.

## 1.5 How this work advances chromatin biophysics

Over time, more studies focus on how the material properties of chromosomes, such as elasticity, alter the architecture of the chromatin fiber. In 2017 a method named ChromEMT found that flexibility changes chromatin compaction (88). With ChromEMT, which couples electron microscopy tomography (EMT) with enhanced DNA contrast, researchers found that disordered chromatin chains have varied flexibility and persistence lengths, resulting in different chromatin compaction levels (88). After publication of my work in Chapters 3 and 4, Shinkai *et al.* used Hi-C data to measure chromosome elasticity computationally (89). Similar to Minimal Cylinder Analysis (MCA), presented in detail in Chapter 4, chromatin's viscoelastic properties were coupled to a thermal bath to interrogate the mechanical response. However, in Shinkai *et al*., elasticity is inferred from Hi-C maps (89), whereas MCA uses high-resolution structural simulations to obtain elastic information.

There are significant sources of error in methods that infer chromatin elasticity from high-throughput experimental data. One source of error is that Hi-C maps are derived from an ensemble average of heterogeneous cells with extensive variance between individual cells (90–92). Consequently, immunofluorescence signals corresponding to certain RNA expression and epigenetic marks vary between individual cells (91). Therefore, one might inadvertently infer flexibility from spatial genomics data due to the artifact of chromatin architectural variability between cells. A potential path to overcome the limitation of ensemble averaging for applications of chromatin dynamics is single-cell methods such as single-cell Hi-C (93), single-cell Assay for Transposase-Accessible Chromatin (scATAC-Seq) (94), or super-resolution microscopy of chromatin (95).

Another source of error in elastic measurements from chromosome conformational capture methods, such as Hi-C, is the inherent multivariance of biological systems. Conclusions drawn from Hi-C ensemble averages may be based on convoluting factors such as dynamic modifications (91) segmenting the chromatin array or active processes. An analogy to explain this is a hypothetical experiment on an origami crane in a box. A camera is placed in a box and captures images of an origami crane with its wings in an array of stances. The wing 'motions' look like flapping; the observer deduces the crane must be flying. An alternative: machinery within the box dynamically altered the fold of the crane. So, is the origami crane just rigid paper bent into varied states, or is it able to smoothly undulate similar to flying? This analogy explains why for models trained on experimental data, it is imperative to

question what precisely is measured. Indeed, numerous active processes distort chromatin such as remodeling (96–98), loop formations from CTFC binding proteins and cohesin (31, 32, 99), and cytoskeletal mechanotransduction (62, 100–102). In the example of the origami crane, and akin to conformation capture methods; the information retrieved is an ensemble of states without information about time evolution. To conclude what the crane does, ideally we would make observations in real time—which is precisely what we can do with molecular dynamics.

An overarching strength of the work presented here is the high-resolution and transparency of simulation setup and data. First, all-atom simulations are primarily physically derived with parameters such as temperature, salt concentration, orientation, and structure tightly regulated. In all-atom MD, minute sub-angstrom resolution fluctuations of macromolecules are detectable. Compared to coarse-grained models, such as AWSEM-MD (74), dynamic changes due to subtle charge alterations are more resolvable. This feature is essential for the work here since the effects of PTMs and structurally similar histone variants (Fig. 1.1) are studied. The ability of all-atom MD to better resolve subtle changes due to mutations may have increased my biological predictions' fidelity.

However, a drawback of all-atom simulations is the issue of sampling and computational cost. Large-scale folding events are generally not accessible, and force probe studies are expensive in all-atom simulations. While atomistic resolution may be necessary for the systems studied here, all-atom simulations also pose a challenge

when measuring material properties. Minimal Cylinder Analysis (MCA) addresses this by calculating the Young's modulus of nucleosomes from structural fluctuations in equilibrium simulations (Chapter 4). To overcome the timescale limitation of our simulation data, computational analysis is coupled to experimental results. My MD results are collaboratively tested *in vitro* and *in vivo* to further contextualize how chromatin's materiomics impact biology.

# Chapter 2: Centromeric dynamics are guided by antagonistic epigenetic marks



Author Contributions: MB and YD designed the biological study; MP, YD, and GAP designed the computational study; MB performed all biochemical and cell biology experiments with assistance from PGD and SR; MP performed all computational experiments; MP, YD, and GAP analyzed the computational experiments; AN and ANL performed and analyzed the MS/MS experiments.

## 2.1 Introduction

Posttranslational modifications in histones play an important role in chromosome biology. The majority of such modifications discovered exist on the N-terminal tails of histones H3, H2A, H3.3, and H4 (1, 103, 104). N-terminal histone modifications may increase nucleosome turnover (105), be inherited at specific loci (106), alter the binding efficiency of various transcriptionally active or repressive factors (107), and disrupt replication timing (108). Thus, modifications can potentially influence the fate of the underlying locus. Research has also uncovered

covalent modifications within histone fold domains, such as H3K56ac and H3K122ac (109–111). H3K122 is acetylated at the nucleosome dyad, wherein it alters DNA–histone binding and increases thermal repositioning of the nucleosome *in vitro* (110). Concurrently, *in vivo* experiments demonstrate that H3K122ac promotes nucleosome turnover, thereby stimulating transcription (112, 113). When mutated, a single residue in the hominid-specific histone variant H3.5, leucine 103, disrupts nucleosome instability both *in vitro* and *in vivo* (114). Further, a single change in the nucleosome, methylation at H3K9, alters replication timing (115). Thus, internal histone core modifications can alter the nucleosome structure in a manner distinct from that reported for histone tail modifications (107, 116, 117). Therefore, investigating how such modifications in key histone variants, such as the centromere-specific H3 histone variant CENP-A, can contribute to function is an exciting area of research. Indeed, it has been previously proposed that specific posttranslational modifications could distinguish newly incorporated from old CENP-A and that new CENP-A not appropriately modified could be evicted during late G1 phase (118). Interestingly, previous work also shows that inhibiting HDACs suppresses the loss of CENP-A at centromeres (119), suggesting that the acetylation of CENP-A plays a role in CENP-A retention.

A plethora of CENP-A modifications have been discovered (120). Of these, only three have been reported within the histone fold domain (78, 121–123). Using epitope-tagged CENP-A, studies have identified phosphorylation of S68 within loop 1 (123), and ubiquitination of K124 near the pseudo-dyad of the nucleosome (121,

122). Bui *et al.* previously analyzed chromatin-bound native CENP-A (nCENP-A) and identified acetylation of K124 (K124ac), which was enriched at G1/S phase (78). K124 in CENP-A is analogous in location to residue K122 in histone H3, which, as discussed above, has a significant impact on nucleosome structure and function. It was also previously reported that in advance of replication, CENP-A K124ac and H4K79ac correlate with a transitionary opening of centromeric chromatin fiber (78). CENP-A K124ac is proximal to the pseudo-dyad DNA of the CENP-A nucleosome and to the CENP-A C-terminus. The latter is required to bind the inner kinetochore protein CENP-C (124). Therefore, we hypothesized that potential functions of CENP-A K124ac/H4K79ac might be to alter the stability of the CENP-A nucleosome. Cumulatively, such events might be necessary to increase access to centromeric chromatin at the appropriate time in replication. Post-replication, inheritance of specific chromatin states involves a coordinated series of events that include eviction of nucleosomes, followed by rapid reassembly after the passage of the replication machinery (125). Consequently, constitutive gain or loss of pre-replicative modifications in CENP-A might be expected to influence centromere replication dynamics.

I dissected the role of CENP-A K124ac *in silico* and formed computational predictions, which were validated *in vivo*. First, using all-atom computational modeling, I simulate the presence of CENP-A K124ac and H4K79ac in the octameric CENP-A nucleosome. I find that these modifications result in a loosening of DNA at the pseudo-dyad, followed by asymmetric site exposure of the terminal ends of the DNA. These DNA dynamics are driven, in part, by an unexpected compaction of the

CENP-A protein, accompanied by a locking of the CENP-A C-terminus. Consistent with this finding, further computational analysis of this acetylated CENP-A nucleosome shows dramatically reduced contacts with the key kinetochore protein CENP-C. To examine the function of CENP-A K124ac *in vivo*, my collaborators generated mutants of CENP-A K124, which mimic either the constitutively acetylated (K124Q), or the unacetylated state (K124A) and we observe that centromere dynamics are impacted. First, gain or loss of CENP-A K124ac results in a quantifiable decrease in association of modified CENP-A nucleosomes with native CENP-C in agreement with computational predictions. Next, relative to wild-type CENP-A, acetyllysine mimic K124Q loses mid–late S phase replication bias, consistent with loosening of DNA discovered *in silico*. A switch was then discovered in modifications of chromatin-bound native CENP-A, from acetylation of K124 at G1/S to monomethylation at S phase. This epigenetic switch may function in part to restore the function of CENP-A to bind CENP-C, which was found to be lost *in silico*.

Together, these results suggest a working model wherein cyclical changes in CENP-A K124 modifications facilitate accurate timing of centromeric replication and contribute to mitotic integrity.

## 2.2 Results

Earlier experimental work from the Dalal lab indicated that G1/S CENP-A nucleosomes are enriched in CENP-A K124ac and H4K79ac, and correlated with a transitionary state of the chromatin fiber (78). To investigate potential dynamics

induced by these modifications, I computationally modeled CENP-A K124ac and H4K79ac in the context of the octameric CENP-A nucleosome. I simulated four systems using all-atom MD: (1) the CENP-A NCP, (2) the acetyl CENP-A NCP (CENP-A K124ac and H4K79ac), (3) the CENP-A NCP bound to the kinetochore protein CENP-C, and (4) the acetyl CENP-A NCP bound to CENP-C. I studied the dynamics and structural time averages of systems 1–4 to determine changes either in proximity of the acetylation sites or even global structural changes. I tested whether these four small charge neutralizations result in detectable perturbations of dynamics in a large system.

### 2.2.1 CENP-A acetylation compacts the histone core

Due to the loss of positive charge on CENP-A K124ac and H4K79ac at the interface of DNA and histones (Fig. 2.1), I hypothesized that DNA contacts with the histone core, and inter-histone repulsion, would both decrease. I tested this hypothesis with detailed contact analysis. Indeed, I found that the acetyl NCP 4-helix bundle interface makes a greater proportion of contacts throughout the simulation (Fig. 2.1a, b). In the acetyl NCP, residues H115, A116, and G117 frequently form more contacts located in the hinge region of the 4-helix bundle. The constraint on this flexible hinge stabilizes the CENP-A to CENP-A′ interface. As a result, the C-terminus—specifically CENP-A′ I132, R133, and G134—forms more contacts with CENP-A G134 (Fig. 2.1b)

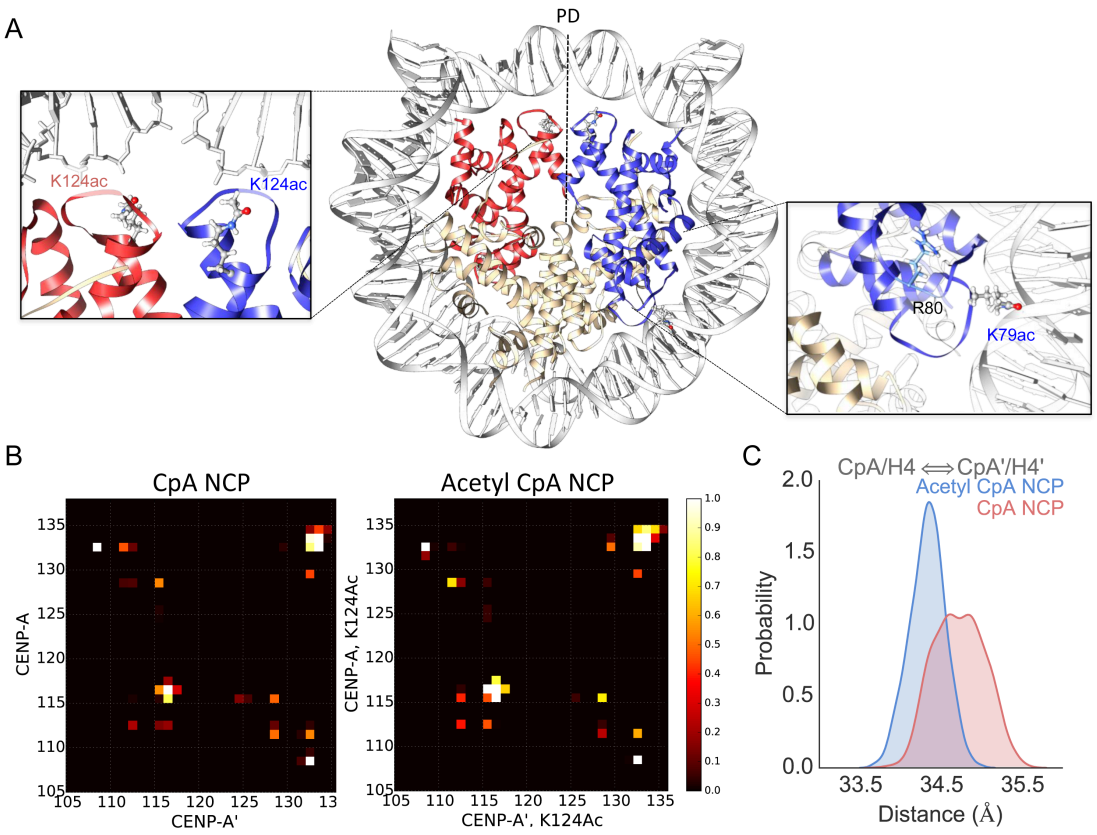


**Figure 2.1: The CENP-A nucleosome displays a tightening of the histone core upon lysine acetylation**. **(a)** The starting structure of the acetyl CENP-A NCP is shown with CENP-A and H4 in *red*, CENP-A′ and H4′ in *blue*, and H2A/H2B monomers in *tan*. The pseudo-dyad is shown by a *dotted black line* and marked as *PD*. Overlay's shown K124ac, K79ac, and CENP-A R80 and the L1 loop in more detail. This structure is after 1-μs simulation of CENP-A, and the production of the starting structure is described (*section 2.5, Methods*). **(b)** Contact analysis showing CENP-A (CpA) to CENP-A′ (CpA′) interface at the 4-helix bundle. The contact cutoff between Cα atoms was set to 8 Å. An increase in contacts is shown upon acetylation—a decrease in the 4-helix bundle interface distance. A value of 1, *white*, shows a contact formed over all simulation time steps and 0, *black*, never. In the acetyl NCP residues H115, A116, and G117 more frequently form contacts located in the hinge region of the 4-helix bundle. The C-terminus, specifically CENP-A′ residues I132, R133, and G134, forms increased contacts with CENP-A G134. (**c**) The center of mass (COM) of dimers was measured over all time steps, and the resulting distributions are shown for CENP-A/H4 to CENP-A′/H4′, the acetylated histones. The acetylated system, shown in *blue*, shows a decrease in variance and distance between the two dimer COMs.

I next wished to assess whether only the histone interfaces tighten in the acetyl CENP-A NCP or whether the globular histone domains also contract. To investigate this, I calculated the distance between the center of mass (COM) of dimers. In results above, CENP-A/H4 and CENP-A′/H4′ dimers are closer together in the acetyl NCP (Fig. 2.1c). In other dimer combinations, such as CENP-A/H4 to H2A/H2B, the variance in the distance between dimers decreased, consistent with the rigidification of the histone core upon acetylation. Overall, the observed changes in distances between various dimers show that the histone core is stabilized and tightened in the acetyl NCP.

To explore these dynamics in greater detail, I next performed principal component analysis of histone core proteins ($PCA^{core}$) where high-frequency vibrational motions are filtered out within the first few PC modes, thereby revealing large global distortions. In this analysis, the most significant mode of motion, $PC1^{core}$, revealed a surprising "freezing" or lack of motion at histone interfaces at the pseudo-dyad, adjacent to the K124ac modifications (Fig. 2.1c; *Appendix A, Supplemental Files,* Movie A1). Movies A1 and A2 are available under 'Supplemental Files' of the ProQuest portal for dissertation storage. Because of the decreased histone rocking in the acetyl NCP, even histone interfaces far away from the studied modifications compact, hence, the effect of acetylated lysines transduce to the other face of the nucleosome (Fig. 2.2a; *Appendix A, Supplemental Files,* Movie A1).

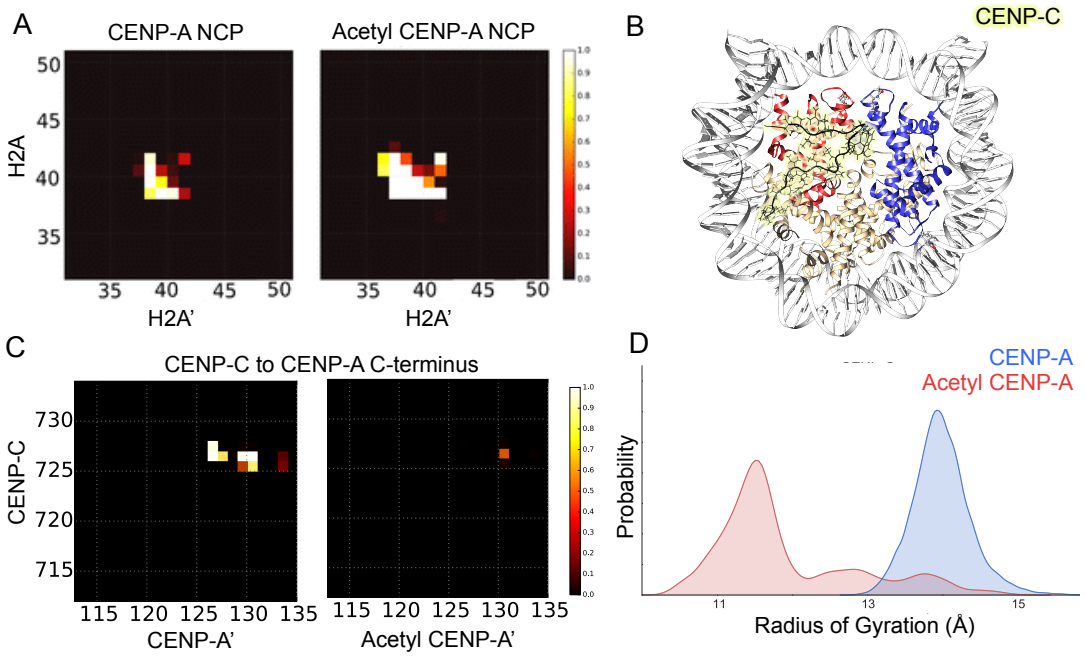


**Figure 2.2: Global compaction of the histone core disrupts CENP-C binding. (a)** Contact analysis of the H2A to H2A′ interface on the opposite face of the nucleosome from the acetylations studied. The contact cutoff between Cα atoms was set to 8 Å. A value of 1, *white*, shows a contact formed over all simulation time steps and 0, *black*, never. An increase in contacts is shown upon acetylation consistent with a global compaction of the histone core transduced away from the points of acetylation. **(b)** Structure of the CENP-C fragment, highlighted in yellow, docked onto the surface of the CENP-A nucleosome with CENP-A/H4 histones in blue and red. **(c)** To study the effects of the histone core compaction with acetylation, the CENP-C fragment (126) was docked to the nucleosomes and contacts analyzed. The CENP-C fragment is shown to form stable contacts with the CENP-A C-terminal end that are lost upon acetylation. **(d)** The radius of gyration of the CENP-C fragment when bound to acetylated CENP-A (red) versus bound to unmodified CENP-A (blue).

### 2.2.2 The CENP-A C-terminus is less accessible in acetyl CENP-A

To compare local structural flexibility, I calculated the root-mean-square fluctuation (RMSF) of Cα atoms (Fig. 2.3a). Compared to the control, the RMSF of the acetyl CENP-A NCP was locally suppressed in specific regions, namely the C-

terminus of CENP-A and the acidic patch of H2A. Both of these regions are targets for CENP-C binding (126). A diminution in the accessibility of the acidic patch may interfere with H2A D89 binding with CENP-C R717 and R719 (127). This unexpected and distinct reduction in the availability of the C-terminus of CENP-A and H2A acidic patch emphasizes local changes that result from these acetylations.

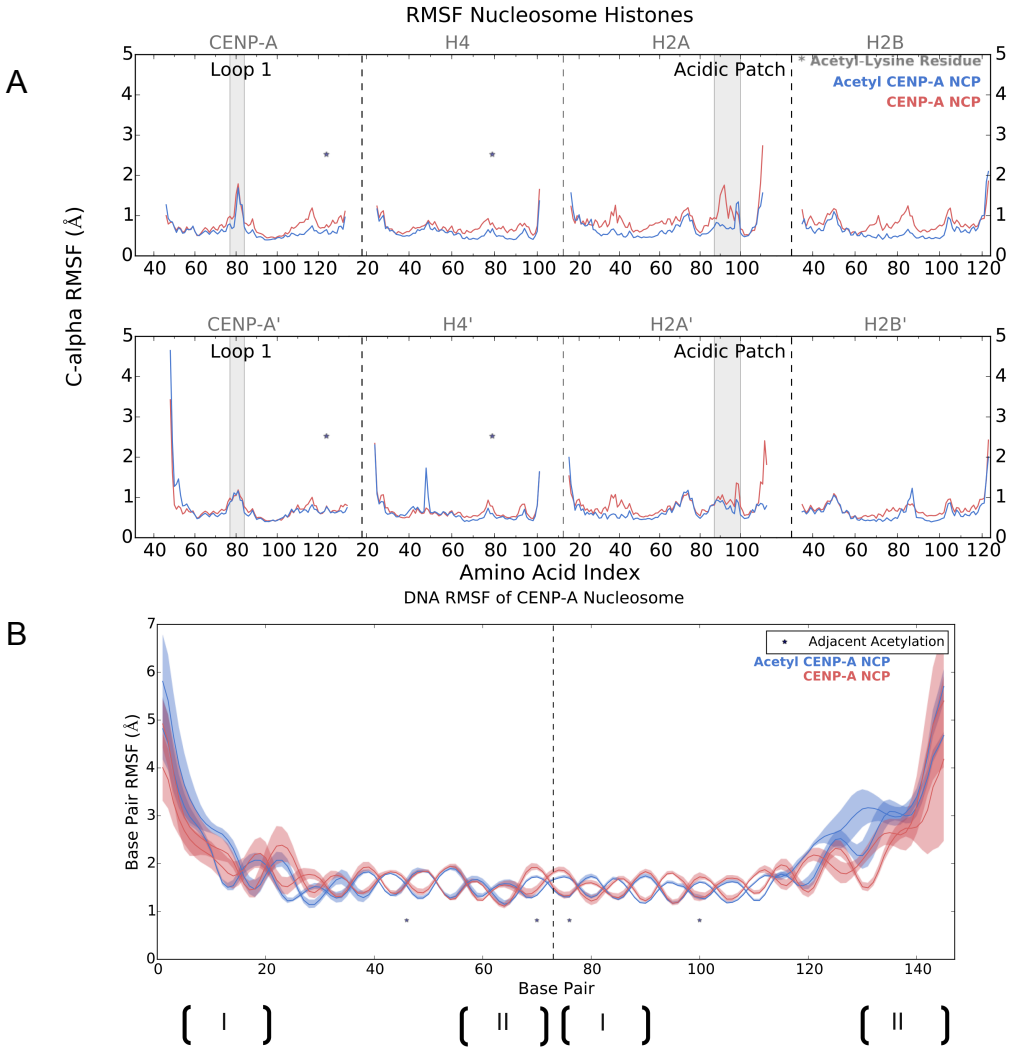


**Figure 2.3: RMSF of proteins and DNA.** (**a**) The decrease in RMSF of Cα residues upon acetylation is more pronounced on the histone heterotetramer adjacent to the entry DNA. Of particular interest, the RMSF of the acetylated H2A acidic patch was suppressed with acetylation by 1 Å, and suppression is shown in the CENP-A C-terminus. The increased similarity shown in the RMSF of the reciprocal histones—CENP-A′, H4′. H2A′, and H2B′—could potentially be explained by the observed asymmetric unwrapping of DNA where the exit end in both systems dissociates a similar amount (Fig. 2.4a). **(b)** The RMSF of nucleic acids is shown for each DNA strand. DNA labled 'I' are DNA wrapped near the DNA entry end or near CENP-A. DNA labeled 'II' are near the exit end of DNA and the histone CENP-A′. The pseudo-dyad is marked by the vertical dotted line.

### 2.2.3 K124, K79 acetylation inhibits CENP-C binding

In order to test whether acetylation modulates CENP-C docking, I docked the CENP-C fragment from the recently solved co-crystal of H3:CENP-A plus CENP-C (126) to each system and simulated for an additional microsecond. A reference for how the CENP-C fragment binds the CENP-A surface is shown in Fig. 2.2b. As can be seen, the CENP-C fragment forms stable contacts with the C-terminus of CENP-A in the unmodified state, but these contacts are virtually lost upon acetylation (Fig. 2.2c). The acetylations modeled result in global structural changes that are sufficient to diminish the accessibility of this critical CENP-C docking interface. Interestingly, in the acetyl NCP, the heterotetramer half showing the greatest suppression of RMSF is also the region with increased DNA unwrapping (Fig. 2.3b, Fig. 2.4a). To further analyze the dynamic motions of the CENP-C fragment, I calculated the radius of gyration of CENP-C bound to the unmodified versus acetyl CENP-A nucleosomes (Fig 2.2d). Expanding upon my finding that contacts are not stably formed between CENP-C and the CENP-A c-terminus (Fig. 2.2c), I find that when CENP-A is acetylated, the CENP-C fragment samples three metastable configurations (Fig 2.2d) and does not remain bound in an extended configuration across the surface as seen in the reference structure (Fig. 2.2b). In summary, my molecular dynamics simulations show that these CENP-A and H4 acetylations lead to a compaction of the histone core, resulting in a relative burial of the CENP-A C-terminus that is normally targeted by CENP-C. As a consequence, CENP-C binding is greatly diminished (Fig. 2.2c). To further study the interaction of histones and DNA, I then extended my analysis to the whole nucleosome.

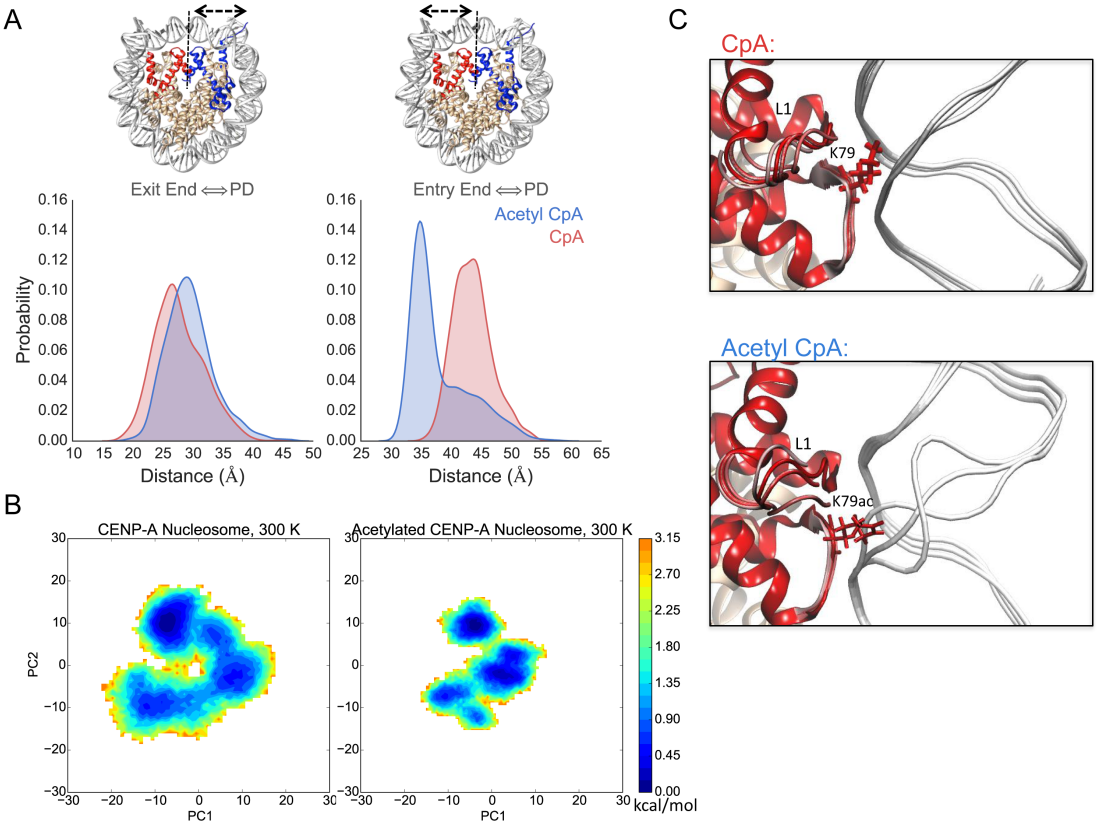


**Figure 2.4: Lysine acetylations asymmetrically loosen DNA entry and exit ends and alter DNA dynamics. (a)** The distance between DNA ends to the pseudo-dyad was measured for all time steps and distributions shown. Structures, on top, show which graph corresponds to which DNA end. In these, the CENP-A NCP has CENP-A and H4 in *red*, CENP-A′ and H4′ in *blue*, and H2A/H2B monomers in tan. The entry end of DNA unwraps more in the acetyl NCP in *blue*. **(b)** Coarse-grained free energy landscapes are shown for CENP-A and acetyl CENP-A from $PC1^{NUC}$ and $PC2^{NUC}$ with whole nucleosomal principal component analysis ($PCA^{NUC}$). Here it is shown that the acetyl NCP landscape becomes more rugged and frustrated. **(c)** From $PC1^{NUC}$ trajectories, three representative snap shots show the intra-helical bubble formed in DNA near K79ac that does not occur in unmodified CENP-A. DNA is shown in *white*, and the L1 loop of CENP-A is marked.

### 2.2.4 DNA is more accessible in the acetyl NCP

In simulation, I observed that the DNA in both systems, CENP-A and CENP-A K124ac, unwraps asymmetrically, but has enhanced unwrapping in the acetyl NCP

(Fig. 2.3b, Fig. 2.4a). I studied the change in DNA dynamics further with principal component analysis of the whole nucleosome ($PCA^{NUC}$). Visual analysis of the most significant mode, $PC1^{NUC}$, demonstrated a pronounced untwisting motion of DNA ends in the acetyl NCP (Movie A2). The rigidification of the histone core in the acetyl NCP stabilizes infrequently sampled states of DNA unwrapping in CENP-A (Fig. 2.4a). I interpret these data to mean that the acetyl NCP system is less freely sampling a larger conformational space, as seen through the attenuation of histone rocking (*Appendix A, Supplemental Files,* Movie A1) and results in a more rugged free energy landscape (Fig. 2.4b).

Another feature exclusive to $PC1^{NUC}$ of the acetyl NCP is a modulation in the widths of the major and minor grooves of DNA (*Appendix A, Supplemental Files,* Movie A2). The cause for this modulation was a pronounced scissoring motion between helices α2 and α3 of the 4-helix bundle in acetyl CENP-A. I observed a high coherence between the scissoring of the 4-helix bundle with the modulation of the size of the DNA minor grooves (correlation coefficient of 0.82, *section 2.5, Methods*). This suggests that the altered motion of the 4-helix bundle in acetyl CENP-A could promote DNA sliding. $PCA^{NUC}$ also revealed that adjacent to both H4 and H4′ K79ac, two intermittent DNA 'bubbles' formed within the double helix (Fig. 2.4c; *Appendix A, Supplemental Files,* Movie A2), indicating that these regions of DNA become more susceptible to opening in the presence of H4K79ac. DNA 'bubbling' can be thought of as melting transition of DNA, where strands locally separate (128). Therefore, the acetylation of H4K79 and CENP-A K124 is sufficient to alter DNA

dynamics and cause asymmetric loosening of DNA ends, inter-helical DNA bubbling adjacent to H4K79, and promotes DNA sliding.

Overall, these all-atom computational modeling results lead to two discrete consequences. First, compared to the CENP-A NCP, the rigidifying of the acetyl NCP locks the CENP-A C-terminus, stabilizes the H2A acidic patch, and disrupts CENP-C access to the CENP-A C-terminus. Second, asymmetric exposure of the DNA coupled with DNA bubbling may promote DNA sliding and thereby open the centromeric chromatin fiber to prime centromeric replication.

To explore these possibilities in a cellular context and to validate my computational predictions, we next turned to an experimental approach *in vivo*. There, my collaborators found that a K124Ac mimic has decreased affinity for CENP-C *in vivo* (*Appendix A,* Fig. A1). This confirms one key computational prediction from my *in silico* modeling (Figs. 2.1-4), that acetyl CENP-A nucleosomes have reduced accessibility at the C-terminus, which results in reduced interactions with CENP-C. Further, my computational findings showed an increased accessibility of DNA in the presence of CENP-A nucleosome acetylation (Fig. 2.4) and that acetyl CENP-A nucleosomes had a higher potential for DNA unwrapping and sliding (Fig. 2.4, *Appendix A, Supplemental files,* Movie A2). Increased unwrapping and sliding could also promote CENP-A nucleosomal mobility. To test whether long-term gain or loss of K124ac impacted centromeric replication, my collaborators studied replication *in vivo* through S phase. We found that CENP-A mutations in K124 display altered replication timing of centromeric foci (*Appendix A,* Fig. A2). This experimental

finding provides further insight into the biological consequence of the dynamic changes to DNA accessibility I found computationally.

### 2.2.5 CENP-A K124Ac switches to monomethylation at S phase

The downregulation of CENP-C binding to CENP-A Ac discovered *in silico* and also seen *in vivo* impedes kinetochore formation at the centromere, necessary for mitosis. Further, the replication defects noted above show that centromeric foci are modestly delayed in replication timing in K124A, and lack a late S phase centromeric replication bias in K124Q cells (Fig. A2). These phenotypes and the necessity for CENP-A to functionally bind CENP-C for mitosis suggest that K124ac must be removed and may occur by early S phase. Deacetylation would allow for centromeric replication and later mitosis to progress correctly. To address whether acetylation at K124ac is removed, or replaced by another modification at S phase, my collaborators utilized a native CENP-A gel-purification strategy coupled to a MS/MS approach (78, 129–131), adding a super-resolving, double long Triton-Acid-Urea (TAU) gel (132). We found that CENP-A K124Ac switches to monomethylation at S phase. These data suggest a cyclical nature to modifications of K124 (*Appendix A,* Fig. A3, A4. The dynamic nature of CENP-A modifications is consistent with computational results. If the dynamic properties of CENP-A Ac, found *in silico*, were sustained through the cell cycle, the centromere would be unable to fulfill it roll for mitotic spindle targeting in mitosis but acetylations are present leading into G1/S to facilitated DNA accessibility as seen *in silico* (Fig. 2.4).

## 2.3 Discussion

Despite decades of intensive biochemical studies, precisely how histone modifications impact biological function has been difficult to dissect. Elegant *in vivo* and *in vitro* experiments demonstrate that core modifications can exert effects on nucleosome behavior, chromatin changes, and biological function (107, 110, 111, 133, 134). In previous work, Bui *et al.* reported that during G1/S, human centromeric chromatin becomes more accessible and that CENP-A nucleosomes occupy a transitionary state concurrent with internal histone fold domain acetylations at CENP-A K124 and H4K79 (78). In this chapter, we probed the effects of CENP-A acetylation using an interdisciplinary approach.

My computational modeling establishes that CENP-A K124ac and H4K79ac collaborate to loosen the DNA at four symmetric contact points within the nucleosome, which is predicted to promote access to the nucleosomal DNA while simultaneously rigidifying the protein core. *In silico* results also suggest that dual acetylation of CENP-A K124 and H4K79 serves an unanticipated role in restraining the C-terminus of CENP-A so that it is less accessible than in the unmodified CENP-A NCP (Figs. 2.1-4). One outcome of this "locking" of the CENP-A C-terminus is a reduced binding to CENP-C *in silico*.

A potential question of interest is how do these acetylations additively alter CENP-A dynamics? The modifications modeled here, H4K79Ac and CENP-A K124Ac, were chosen to dissect their role in centromere regulation during G1/S of the cell cycle leading into DNA replication. While the specific lysine acetylations

modeled here were discovered to be enriched before replication and may be in a fully modified state (135), alternative modification profiles could exist either transiently or as a stable state within the centromere. For example, CENP-A nucleosomes could be acetylated asymmetrically across the pseudo-dyad. Further, while there is a reduction in number of states due to the pseudo-symmetry of the nucleosome, there are nine distinguishable combinations of PTM states when only counting for H4K79 and CENP-A K124 acetylations (all four unacetylated, one copy of K79Ac, two copies of K79Ac … all four acetylated). The precise dynamic alterations due to asymmetric acetylations or any of the seven other unstudied modification patterns are unknown. It was later shown that combinations of histone core charge modifications can be unpredictable in resulting DNA affinity and are nonlinearly constructive or destructive in effect (136). Therefore, further dynamic studies would be required to understand additional combinatorial effects of the PTMs studied here. Nucleosome modification heterogeneity in the nucleosome modifications is a variable controlled here, *in silico*, but could be pertinent in native CENP-A studies *in vivo* (Figs. 2.5-6).

In testing the hypothesis that CENP-C binding is altered, we noticed that both K124Ac mimics have reduced CENP-C binding (*Appendix A,* Fig. A1). Decompaction of the CENP-A chromatin fiber may be a prerequisite to replication. *In vitro* studies of unmodified CENP-A on 601 DNA sequences (137, 138) have reported that CENP-A chromatin fibers are more compacted. We speculate that G1/S-specific CENP-A K124ac/H4K79ac promotes unfolding of the centromeric fiber, possibly by increased nucleosomal sliding as seen *in silico*, and transient release of CENP-C from the CENP-A C-terminus, repositioning CENP-C to the intervening

linker DNA. Recent work shows that centromeric replication timing is exquisitely sensitive to the state of the chromatin fiber. Importantly, the both the computation and experimental results point to the importance of *removing* the acetylation. Here, we found that CENP-A K124 acetylation is replaced by monomethylation at S phase (*Appendix A,* Fig. A4).

Precedence for histone core domain methylation in regulating replication also exists. Methylation of H3K79 limits genome-wide DNA replication to once per cell cycle, thereby preventing over-replication (139). We hypothesize that one putative function of K124me may be to stabilize the nucleosomal DNA, either by rapid re-binding of CENP-C immediately after replication, or by inhibiting rebinding of pre-RC components to the newly replicated CENP-A chromatin. Additional future experiments that are necessary include dissecting how HATs like p300, which contributes to CENP-A′s acetylation (40), are recruited to the centromeric fiber, and identifying the HMT responsible for CENP-A K124 methylation. Excitingly, ongoing *in vitro* dissections using synthetically engineered CENP-A K124ac (140) coupled to H4K79ac proteins, will determine whether acetylated CENP-A nucleosomes encode a chromatin fiber more susceptible to sliding as indicated from our *in silico* data, and whether such a fiber has an altered affinity for CENP-C as discovered *in silico*.

## 2.4 Conclusions

I dissected the role of modifications in the CENP-A nucleosome *in silico* and my results and predictions were seen *in vivo*, where these acetylations are shown to

be involved in regulating CENP-C distribution on the modified CENP-A chromatin fiber, mitotic integrity, and centromere replication timing. Further, this work provides physical insight as to why CENP-A must have cyclical posttranslational modifications (K124ac → K124me), so that mitosis is able to proceed. We emphasize that the delays in centromere replication timing observed are not permanent, as most cells recover by late S phase. Thus, while acetylation of CENP-A K124 does not appear to play a deterministic role, it does subtly contribute to centromere dynamics at the level of replication timing and mitotic integrity. Consequently, these data open a new avenue of investigation into how covalent modifications, buried within the histone fold domain of histone variant nucleosomes, can serve as epigenetic regulators of biological processes.

## 2.5 Methods

### 2.5.1 Simulation protocol

All-atom molecular dynamics (MD) simulations were performed with software suite Gromacs 5.0.4 (141). The force field employed to model nucleosomes was amber99SB*-ILDN (142, 143) for proteins, amber99SB parmbsc0 (144) for DNA, ions08 (145) for ions, and the TIP3P (146) water model.

Two nucleosomal systems were built for simulation: the acetyl-lysine CENP-A nucleosome and then unmodified CENP-A nucleosome as control. First, the CENP-A nucleosome was built with PDB ID: 3AN2(15)—resolved to 3.60 Å—as the starting structure. Unresolved 3AN2 residues Thr 79 through Asp 83 of CENP-A′, Chain E, were built with MODELLER (147). During energy minimization of this

constructed region, one residue in the n-terminus and c-terminus directions was unconstrained. Additionally, heterogen selenomethionine residues were altered to methionine through a single-atom mutation from Se to S. As a control, the 146 base-pair α-satellite DNA of PDB ID: 3WTP(21) was aligned onto 3AN2 using the CE algorithm (148) of PyMOL(149). To further study the effect of acetylation on CENP-C binding, the CENP-C fragment(126) was docked onto each system using the CE algorithm(148) and simulated for an additional microsecond retaining identical simulation protocols. This is the first known all-atom detail structural analysis of CENP-C bound to CENP-A.

From this initial structure, the Gromacs tool pdb2gmx was used to assign charges to residues at biological pH: a charge of +1 on Lys and Arg, 0for Gln, −1 for Asp and Glu, and His is neutral with the ε-nitrogen protonated. Then, a rectangular cuboid box was created such that boundaries were a minimum distance of 1.5 nm from the unsolvated system. Next, Na+ and Cl− ions were introduced to neutralize the system charge and additionally model an ionic concentration to 150 mM. For both pre-production and production runs, periodic boundary conditions were employed. Electrostatics were handled with the particle-mesh Ewald method and Verlet cutoff scheme. For the non-bonded interaction shift functions, Coulombic and van der Waals potentials had a cutoff distance at 1.0 nm. Covalent bonds to hydrogen were constrained with the LINCS algorithm.

The CENP-A nucleosome system was energy minimized using steepest descent to a maximum energy of 100 kJ/mol. The systems were then equilibrated in

multiple steps. First, the systems were heated to 300 K for 2000 ps. During this step, DNA was restrained with $K = 1000$ kJ mol$^{-1}$ nm$^{-2}$ in the canonical ensemble (NVT). For the next thermal equilibration at 300 K for 2000 ps, both DNA and protein had weak harmonic position restraints, $K_{\text{CENP-A}} = 2.5e{-}5$ kJ mol$^{-1}$ nm$^{-2}$, to hinder global rotational motions (explained in 2.5.2). Lastly, pressure was equilibrated for 1500 ps in the isothermal isobaric, NPT, ensemble at 300 K and 1.0 bar with $K_{\text{CENP-A}}$.

This system was ran for 1 μs at 300 K. Temperatures were V-rescaled which is velocity rescaling with a stochastic term to achieve the proper ensemble (150) with a time constant of 1.0 ps. System pressures were regulated with the Parrinello–Rahman barostat(151) at 1.0 bar and a time constant of 2.0 ps. The simulations' time step size was 2 fs. Coordinates, velocities, and energies were saved every 2 ps. Non-bonded neighbors lists were updated every 20 fs.

After the CENP-A nucleosome was run for 1 μs, the final structure was acetylated in four histone core locations: K124 of CENP-A and CENP-A′, and K79 of H4 and H4′. The partial charges assigned to acetyl-lysine atoms were calculated quantum mechanically as described previously(152). The new amino acid type for acetyl-lysine, KAC, was added to amber99SB*-ILDN[2,3]. Both the acetyl-lysine CENP-A system and the control CENP-A system were simulated for an additional 1 μs as described above. For subsequent analysis, trajectories were truncated to remove the first 600 ns to account for additional system equilibration during production runs.

### 2.5.2 Position restraints

Periodic boundary conditions (PBCs) are used so that there are a decreased number of solvent atoms without experiencing artificial effects from simulation boundary interactions. However, for PBCs to be judiciously employed, the macromolecules in each virtual repeating unit cell must not come within a cutoff distance where the simulated structure self interacts across the boundary. For large N all-atom simulations where long time scales are desired, there are computational benefits to decreasing the simulation box size.

Likewise, the distance between virtual images may decrease if the system rotates. With this in mind, I employed a position restraint on each atom in the simulation to attenuate rotational effects while holding the simulation box at a more minimal size. I developed a brief analytical model to calculate the position restraints of nucleosomes.

The toy model for the weak position restraint calculation of nucleosomes uses a freely rotating rigid rod model. Alternative geometries could be employed. One could imagine that the nucleosome in a periodic box is modeled as a heterogeneous rod formed from a circle's projection onto its diameter to form a rod. By using a heterogeneous rod, the mass density depends on cross-section width along the axis of projection. I then calculate a harmonic potential acting on points along the rod to inhibit rotations greater than $\theta_{max}$. As an example, two arbitrary points are selected: point R, which is the edge of the rod furthest from the axis of rotation, and point $y_2$

located some arbitrary distance along the rod, y, from the axis of rotation at the center of mass:

$$\varepsilon_R = \frac{1}{2} k \Delta x_{max}^2 = \frac{1}{2} k R^2 4 \sin^2\left(\frac{\theta_{max}}{2}\right), \quad (2.1)$$

$$\varepsilon_{y_2} = \frac{1}{2} k \Delta x_2^2 = \frac{1}{2} k y_2^2 \, 4 \sin^2\left(\frac{\theta_{max}}{2}\right), \quad (2.2)$$

$$\varepsilon_{TOT} = \sum_{i=0}^{R} \varepsilon_i \, N(y_i)$$

where $\varepsilon$ is the harmonic potential, $N(y_i)$ is the number of atoms at distance y, k is the spring constant, and x is the displacement. Instead of considering a rod comprised of discrete points, I will treat it as a continuum and integrate over the rod. I assume constant density on the face of the circle:

$$\delta \approx \frac{N}{\pi R^2}$$

where $\delta$ is the density, N is the number of atoms, and R is the radius. I will then integrate over the distribution of N(y):

$$\varepsilon_{TOT} = \frac{1}{2} k \int_{-R}^{R} N(y) \Delta x^2 dy. \quad (2.3)$$

Factoring out the constant factors of $\Delta x$ from Eqs. 2.1 and 2.2, plugging in the density of both faces, and solving for Eq. 2.3 in terms of y and R:

$$\varepsilon_{TOT} = 2k \sin^2\left(\frac{\theta_{max}}{2}\right) \int_{-R}^{R} \frac{2N}{\pi R^2} y^2 \sqrt{R^2 - y^2} dy,$$

$$C = \frac{4kN}{\pi} \sin^2\left(\frac{\theta_{max}}{2}\right),$$

$$\varepsilon_{TOT} = C\frac{1}{R^2}\int_{-R}^{R} Ry^2 \sqrt{1-\left(\frac{y}{R}\right)^2}\, dy.$$

By substitution, if y/R = sin(x) and dy = Rcos(x)dx. The integral can then be simplified to

$$\varepsilon_{TOT} = CR^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2 x \cos^2 x \, dx = CR^2 \frac{\pi}{8}.$$

Plugging back in the value of C and giving 1/2 $k_b$T energy to rotation, where T is temperature and $k_b$ is Boltzmann's constant, I solve for the spring constant to find:

$$k = \frac{k_b T}{NR^2 \sin^2\left(\frac{\theta_{max}}{2}\right)} \quad .(2.4)$$

System parameters are the distance between nucleosome images, in this case about 30 Å apart; the number of atoms; and the radius. I restrict the nucleosomes from rotations where any surface comes within 10 Å. Therefore, $\theta_{max} = \arccos\left(\frac{10\text{ Å}}{30\text{ Å}}\right)$. From Eq. 2.4 the position restraints are calculated.

### 2.5.3 Analysis of trajectories

After truncating the simulation data to the final 400 ns for analysis, the root-mean-square fluctuations (RMSF) were calculated for the Cα atoms of the histones

and the average of the nucleic acids for DNA. The RMSF is used to calculate local time-averaged fluctuations. The RMSF of DNA (Fig. 2.3b) was calculated for thirds of the final 400 ns and then the standard deviation of the mean plotted. Contact analysis was calculated with a cutoff distance of 8 Å between histone Cα atoms to compare dimer interface contacts in both systems. The center of mass (COM) of dimers was then calculated along the trajectory and the distribution of distances between COMs compared.

Principal component analysis (PCA) was performed on the histone core as previously described[2]. This analysis was then extended to include Cα atoms of histones and phosphate atoms nucleosomal DNA. The first and last ten base pairs were truncated from the analysis to remove ends calculated to have a high RMSF (Fig. 2.3b). This alteration was made so that DNA end motions did not dominate the major principal components. The magnitude of motion is multiplied by a factor of 5 in the movies to amplify motions for visual clarity.

Experimental methods for the data discussed within this chapter are available in *Appendix A*. Extended experimental methods are available in the published version of this work (36).

# Chapter 3: Intrinsic elasticity of nucleosomes is encoded by histone variants and calibrated by their binding partners

*This chapter is adapted from: Daniël P. Melters*, Mary Pitman*, Tatini Rakshit*, Emilios K. Dimitriadis, Minh Bui, Garegin A. Papoian, and Yamini Dalal. "Intrinsic elasticity of nucleosomes is encoded by histone variants and calibrated by their binding partners". Proceedings of the National Academy of Sciences, 116 (48) 24066-24074. (2019). *Co-first authors* (73)

Author contributions: DPM, MP, TR, EKD, GAP, and YD designed research; MP performed computational research, DPM, TR, EKD, and MB performed experimental research; DPM, MP, TR, EKD, and YD contributed new reagents/analytic tools; DPM, MP, TR, EKD, MB, GAP, and YD analyzed data; and DPM, MP, GAP, and YD wrote the paper.

## 3.1 Introduction

The adaptive nature of chromatin allows a cell to replicate, divide, differentiate, regulate transcription, and repair damaged DNA. In part, the chromatin landscape is shaped by removing old and incorporating new nucleosomes with specific histone variants, and by incorporating covalent modifications (9). How different histone variants convey the unique mechanical properties of their nucleosomes to the chromatin fiber, and whether non-canonical nucleosomes modulate chromatin dynamics is a subject of intense study. In contrast to the previous view that chromatin was a mostly static packaging polymer, several studies have

unveiled a rich conformational landscape of nucleosomes (1–6, 153, 154). These works raise the intriguing possibility that mechanical properties embedded within evolutionarily distinct nucleosome types might lead to different structural outcomes for the chromatin fiber. Indeed, recent computational modeling has linked specific epigenetic chromatin modifications to chromosome architecture, genome folding, and genome dynamics (153). Paradoxically, the most evolutionarily divergent histone variant between species are CENP-A orthologs (155), which is functionally essential across most eukaryotes (34, 99, 156). Another major paradox is that despite being buried in pericentric heterochromatin (19), CENP-A chromatin is transcriptionally active in most species, suggesting this chromatin is accessible even when bound to kinetochore proteins (155, 157, 158). This puzzling dichotomy can be explained either by intrinsic mechanical properties, or by epigenetic alterations driven by chromatin effectors.

To investigate this salient problem, I analyzed how binding partners tune the dynamic properties of centromeric nucleosomes. Further, I developed *in silico* tools to dissect innate mechanical properties of CENP-A nucleosomes relative to their canonical counterparts, in the presence or absence of CENP-A binding partners. My findings were quantitatively verifies *in vitro* and a potential biological model put forward. We find that the smallest unit of the chromatin fiber can have profound effects on the three dimensional folded properties of chromatin, with implications for the accessibility of chromatin to transcriptional machinery.

## 3.2 Results

### 3.2.1 CENP-$C^{CM}$ increases the Young's modulus of CENP-A *in silico*

I first examined elasticity as a mechanical feature of nucleoprotein complexes, which had not been reported before. Using all-atom molecular dynamics, I measured nucleosome stiffness and examined spontaneous structural distortions that occur in the presence of CENP-C. I ran three simulations for this study: (1) the CENP-A nucleosome core particle (NCP), (2) the CENP-A NCP with one bound rat CENP-C motif of CENP-C (CENP-$C^{CM}$), and (3) the CENP-A NCP with two copies of CENP-$C^{CM}$. As a control, I compared these systems to canonical nucleosomes, H3 (159, 160).

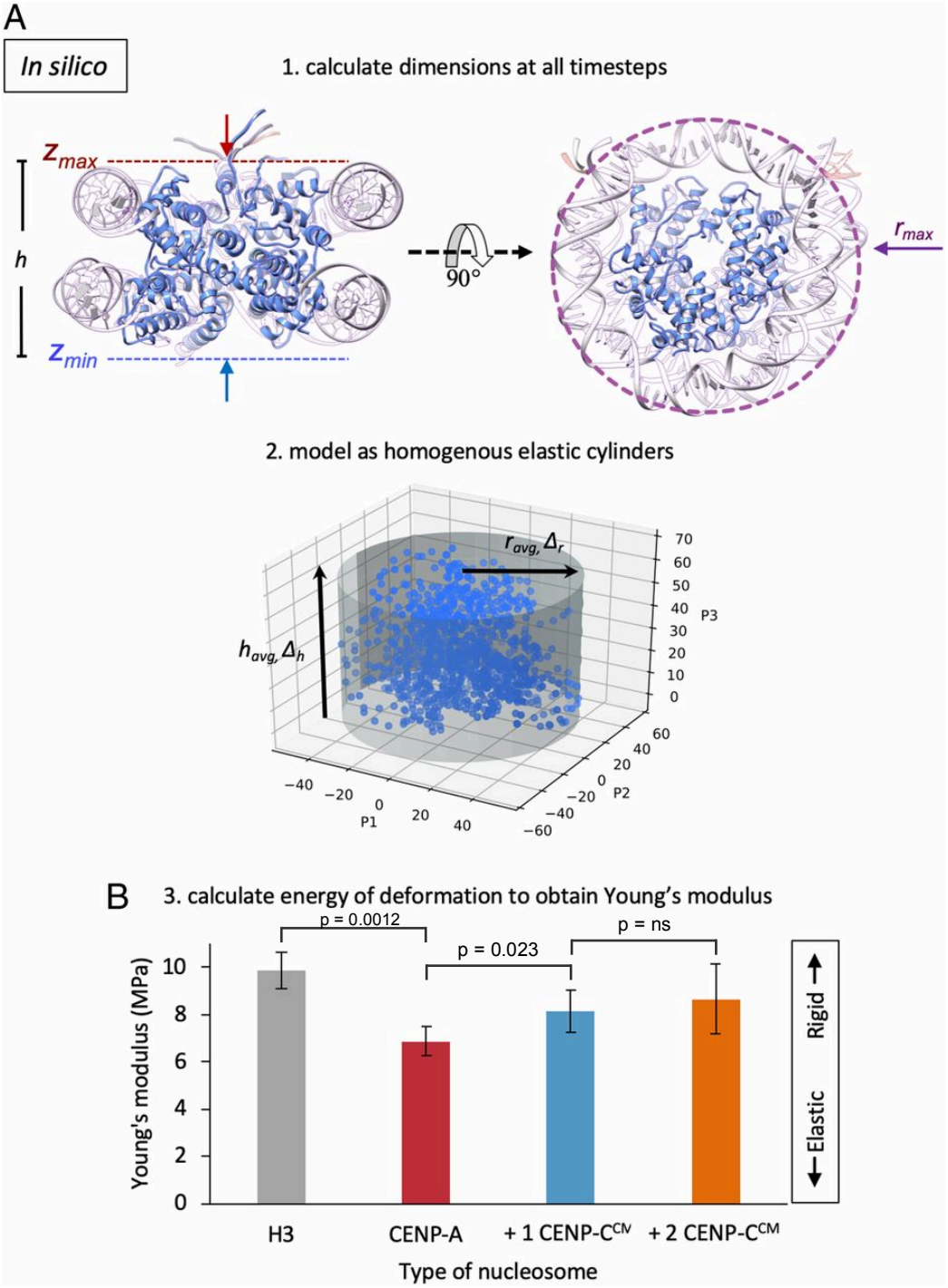


**Figure 3.1: *In silico* analysis predicts that CENP-A nucleosomes are more elastic than H3 nucleosomes. (a)** To obtain Young's modulus values from simulation, I measured the *in silico* dimensions of nucleosomes by compression of an encapsulating cylinder programmed to stop at stiffer surfaces resistant to collapse. From the heights, $h = z_{max} - z_{min}$, and the radii, $r_{max}$, of the resulting minimal

cylinders I then calculated the average and change in height ($h_{avg,}$ $\Delta h$), and radius ($r_{avg,}$ $\Delta r$) of each system. **(b)** I treated the nucleosomes as elastic homogenous cylinders, calculated the energy of deformation, and retrieved the Young's modulus of a cylinder vibrating at equilibrium in a thermal bath. P-values are calculated for an alpha value of 0.05.

Using these all-atom data, I next developed a novel analytical technique to quantify the elasticity of nucleosomes *in silico*. Briefly, this technique connects structural fluctuations observed in unbiased molecular dynamics simulations, with the nucleosome's mechanical response, ultimately producing the absolute value of the Young's modulus (*section 3.4, Methods*). For an extended presentation of an adapted version of MCA, see Chapter 4. To analyze all-atom simulation data in such a way, I modeled the nucleosomes as mechanically homogenous elastic cylinders vibrating in a thermal bath and calculated the dimensions and fluctuations of these "minimal" cylinders during each simulation trajectory (Fig. 3.1a). I further visualized the differences in fluctuations among different types of nucleosomes, finding a distinct height difference of the nucleosome core particle when bound to CENP-$C^{CM}$ and an overall collapse in the variance of fluctuations (Fig. 3.2a-d). It is important to note that for the elasticity measurements of CENP-A + CENP-$C^{CM}$ residues of CENP-C were not included in the analysis set. Therefore, the height differences found for the average heights of bound and unbound CENP-A (Fig. 3.2d) are purely due to a structural transition of the CENP-A nucleosome when bound to CENP-C.

These analyses predict that the Young's modulus of CENP-A is noticeably more elastic (6.2 ± 0.6 MPa) than that of H3 (9.8 ± 0.8 MPa). Interestingly, upon binding either one or two CENP-$C^{CM}$ fragments (Fig. 3.1b), CENP-A nucleosomes'

Young's moduli increase and broaden (8.2 ± 0.9 MPa and 8.7 ± 1.5 MPa, respectively), nudging their upper range of the elasticity profiles closer to that of H3. These findings speak to the elasticity of the multimeric nucleosome structure in comparison to its constituent parts. Indeed, in DNA stretching experiments, where DNA was pulled laterally, the Young's modulus was found to be 3.3 GPa, while from a rod-bending model, 300 MPa was suggested (80). On the other hand, unrelated multimeric protein complexes, such as an antibody pentamer, were found to have Young moduli of 2.5 – 9 MPa (161), approximately in the same range as for the nucleosome core particles reported in this work.

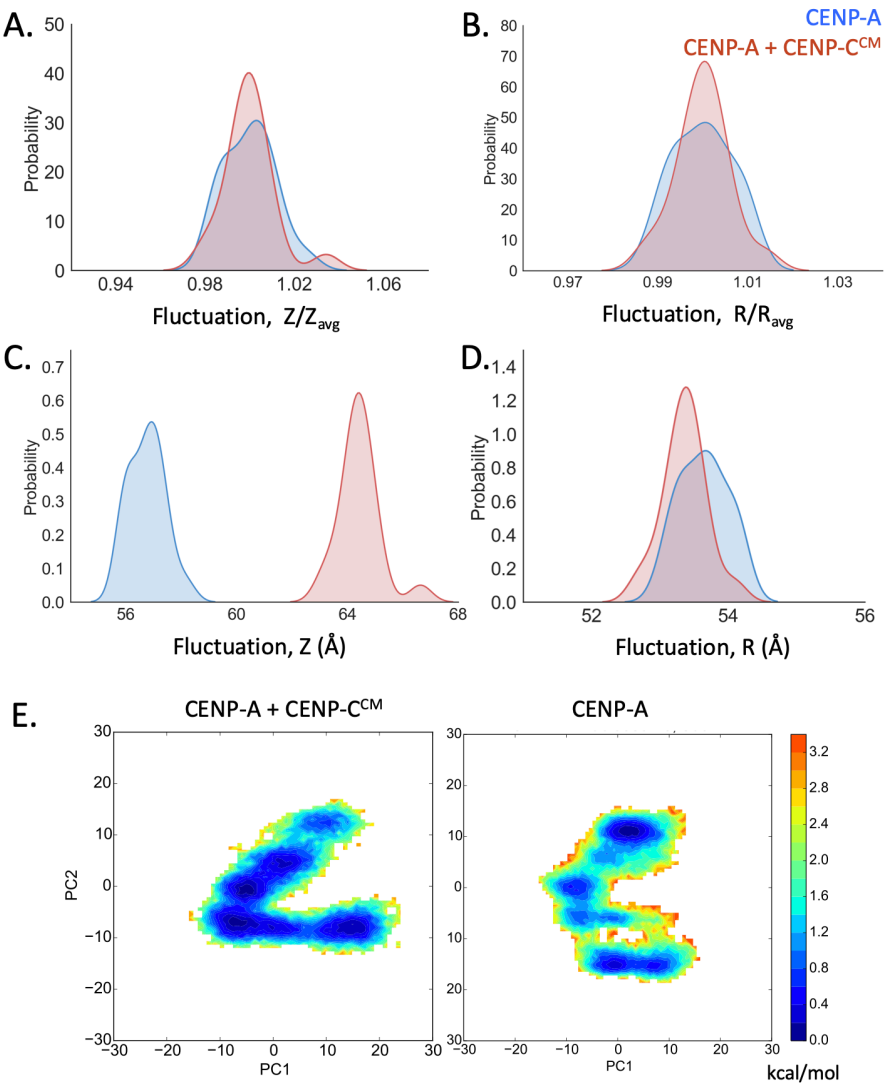


**Figure 3.2: CENP-C$^{CM}$ limits CENP-A nucleosomal fluctuations.** Raw data of the calculation of Young's modulus are shown for CENP-A and CENP-A + CENP-C$^{CM}$ for comparison. CENP-A is shown in blue and CENP-A + CENP-C$^{CM}$ in red. **(a)** The ratio of the measured height at each simulation segment to the average nucleosome height is shown. **(b)** Raw data for radial distributions is shown. **(c)** Height measurements are shown and reveal a distinct difference in the heights of the nucleosome core particles. The CENP-C$^{CM}$ fragment is excluded from height measurements. **(d)** Radial measurements show increased fluctuation for CENP-A but with similar dimensions. The fluctuations are calculated from the variance of plots **c** and **d** and the Young's modulus derived from the strain energy density is sensitive to volume, such that the same magnitude fluctuation on a larger structure obtained increased values of Young's modulus. **(e)** The sampling of the first two principal components (PC1 and PC2) are shown to obtain coarse grained free energy landscapes. Free energy minima are shown in blue and less frequently sampled conformations correspond to red. CENP-A shows a glassier free energy landscape with higher energetic barriers between minima compared to CENP-A + CENP-C$^{CM}$.

### 3.2.2 CENP-C interactions suppress spontaneous structural distortions of CENP-A nucleosomes

The above discussed variation of elasticity of distinct nucleosomal moieties made us curious to examine conformational changes of CENP-A mononucleosomes that might be induced by CENP-$C^{CM}$. To characterize the global motions of these complexes, I carried out Principal Component Analysis (PCA), which is a method to identify larger amplitude and slower frequency motions ranked by variance (162). These modes, which are akin to normal modes when analyzing molecular vibrations, are called the Principal Components (PCs). Subsequently, I obtained PCA free energy (FE) plots through the histogramming of the first two PCs (Fig. 3.2e). I found a somewhat rough FE landscape of the CENP-A nucleosome, similar to previous studies on the origins of CENP-A's intrinsic motions (163, 164). However, upon binding of the CENP-$C^{CM}$ fragment, which associates across the exterior of the histone core, the FE minima coalesce, showing a reduction in metastable configurations (Fig. 3.2e). Furthermore, PCA revealed a dampening of histone motions relative to each other upon binding of CENP-$C^{CM}$ (Fig. 3.2e), which is consistent with the above discussed elasticity observations.

I was next curious to assess how these changes would propagate through the DNA. Thus, I investigated DNA gyre sliding and gaping of nucleic acids through *in silico* labeling (Fig. 3.3a). Indeed, a single CENP-$C^{CM}$ fragment dampens CENP-A nucleosome gyre gaping and DNA slides asymmetrically away from the CENP-$C^{CM}$

bound-face of CENP-A nucleosomes (Fig. 3.3a). I performed additional structural analysis to demonstrate local structural elasticity. Altogether, detailed analyses of CENP-A mononucleosomes motions revealed a global dampening of innate motions upon CENP-$C^{CM}$ binding (Fig. 3.3b). On the residue scale, I found that CENP-$C^{CM}$ suppresses residue fluctuations with symmetry breaking in the presence of one fragment (Fig. 3.3c). These computational data are in agreement with experimental observations made by sm-FRET and hydrogen/deuterium-exchange mass-spectrometry (80) for the CENP-A nucleosome bound to the central domain region of human CENP-C (CENP-$C^{CD}$). The CENP-$C^{CM}$ and CENP-$C^{CD}$ bind to CENP-A nucleosomes through the same mechanisms (165–167), likely because both domains contain the H2A/H2B acid patch binding motif (RR(S/T)nR) and the CENP-A C-terminal tail binding residues (WW/YW), which are separated by seven residues. Importantly, these two motifs in CENP-C are conserved across plant, fungi, and animal kingdoms (9). These data predict that CENP-C dampens motions of CENP-A nucleosomes, and as a consequence, alters mechanical properties of the CENP-A nucleosome.

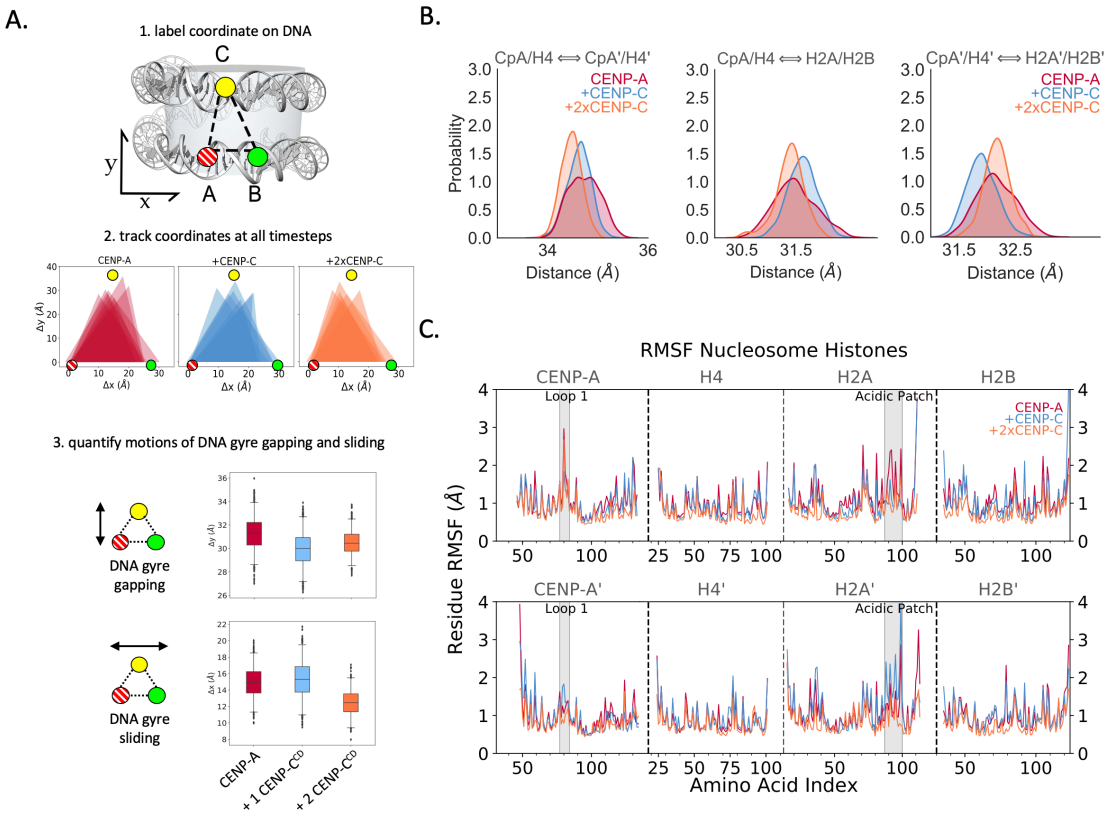


**Figure 3.3: Two CENP-C$^{CM}$ fragments strengthen stiffening of CENP-A nucleosomes. (a)** All-atom computational modeling of DNA gyre gapping or DNA gyre sliding of CENP-A nucleosome alone or bound to either 1 or 2 CENP-C$^{CM}$ fragments. **(b)** The distance between the center of mass (COM) of CENP-A/H4 dimers is shown in red for CENP-A, blue for CENP-A + 1 CENP-C$^{CM}$, and in orang for CENP-A + 2 CENP-C$^{CM}$. Two CENP-C$^{CM}$ fragment exaggerated the COM distances compared to a single CENP-C$^{CM}$ fragment, which means that 2 CENP-C$^{CM}$ further induces a global loss of CENP-A nucleosome elasticity. **(c)** Residue root mean square fluctuations (RMSF) shows freezing of local fluctuations in the CENP-A nucleosome shown in red, 1 CENP-C$^{CM}$ bound shown in blue, and 2 CENP-C$^{CM}$ bound shown in orange. In the region of CENP-C$^{CM}$ binding, the first heterotypic half on the top panel, CENP-C is seen to freeze the acidic patch and the loop 1 region of CENP-A. One CENP-C$^{CM}$ creates asymmetry, especially at the C-terminal end of H2A and H2B, this asymmetry is reduced when the second CENP-C$^{CM}$ is bound. Dashed lines separate individual histones.

### 3.2.3 Experimental findings agree with *in silico* results

To experimentally test my computational calculations, my collaborators used nanomechanical force spectroscopy (126) of reconstituted chromatin (168–174) *in vitro*. This single-molecule method is used to physically compress and release complexes to directly quantify their elasticity on a nanoscale (175, 176). First, consistent with the computational model I developed to mimic AFM computationally (MCA), CENP-A and H3 mononucleosomes display uniform elasticity across their surfaces, behaving as homogenous cylinders (*Appendix B,* Fig. B1). Remarkably consistent with my computational results (Fig. 3.1b), and with the result for mononucleosomes (9), the effective Young's moduli of H3 and CENP-A nucleosomes are distinct. The Young's modulus of H3 nucleosomes is 11.3±4.1 MPa, whereas CENP-A nucleosomes are nearly twice as elastic, at 5.8±3.0 MPa (*Appendix B,* Fig. B2b).

Further, my *in silico* experiments predicted that CENP-$C^{CM}$ suppresses CENP-A nucleosomal motions and consequently innate elasticity (Fig. 3.1). My collaborators tested this prediction *in vitro*. My data show that *in silico,* CENP-A nucleosomes possess innate elasticity and that CENP-C effectively suppresses the freedom of motions of CENP-A nucleosomes. *In vitro,* when CENP-C is bound to CENP-A at both 2 and 4-fold excess a loss of CENP-A elasticity is also found (9). From a thermodynamic perspective, I hypothesized that elastic particles possess higher configurational entropy (30, 79). In other words, elastic particles tend to be less ordered. Thus, we were curious to test whether nucleosomes with a broadened

range of configurational states might collectively form less ordered chromatin and energetically disfavor compaction. Indeed, increased clustering of chromatin arrays was detected (9). My experimental collaborator found that overexpression of CENP-C, which was found to compact chromatin, correlates with decreased RNAP2 loading (*Appendix B,* Fig. B3).

## 3.3 Discussion

Not all nucleosomes are identical, as many contain histone variants, giving them distinct structures and functions (177). We systematically teased apart how a single histone variant encodes mechanical properties to its nucleosome, which were dramatically modified by a small fragment of its cognate protein partner. Using novel *in silico* computational modeling and *in vitro* single-molecule nanoindentation force spectroscopy, we directly measured effective elasticity of nucleosomes and found that CENP-A is more elastic than canonical nucleosomes (Fig. 3.1-3). Indeed, we found remarkable agreement between the computation model to derive the Young's modulus, and the experimental data measuring the elasticity. Second, our findings of noticeably elastic CENP-A nucleosomes have important implications. On the one hand, softer CENP-A nucleosomes are expected to undergo more vigorous structural fluctuations, in turn, potentially exposing cryptic binding surfaces that may facilitate various association outcomes. On the other hand, softer CENP-A nucleosomes may contain excess entropy compared to canonical nucleosomes, which, in turn, would suggest an additional entropy loss upon formation of compacted CENP-A chromatin. However, dynamics of different histone variants are difficult to predict *a priori*: for

instance, H3 nucleosomes within HP1 chromatin show, surprisingly, increased dynamic behavior (2, 6, 153). Therefore, although we anticipate additional entropic resistance to compaction for chromatin enriched with CENP-A nucleosomes, it will be exciting to apply tools developed in this work in future studies for other important types of nucleosomal complexes.

CENP-C is the essential CENP-A binding protein, which facilitates the assembly of the kinetochore (178, 179), and has been shown to alter local CENP-A nucleosomes dynamics (180–182). Previous FRET and hydrogen/deuterium exchange mass spectrometry experiments focused on how CENP-$C^{CD}$ binding alters internal CENP-A mononucleosome dynamics. These data show that human CENP-$C^{CD}$ restricts DNA gyre gaping, sliding, and protects the internal H4/H2A interface (165–167). In prior computational modeling Winogradoff *et al.* showed that CENP-A nucleosomes sample broadened conformational states (165–167). From this, we predicted that CENP-C limits configurations of CENP-A nucleosomes. Indeed, when we modeled CENP-A nucleosomes alone, vs. those bound to CENP-$C^{CM}$, I observed marked diminution of nucleosome motions, and increased Young's moduli, representing lost conformational elasticity (Fig. 3.1b, Fig. 3.3). A physical analogy is that CENP-C behaves as a nanoscale staple on the surface of the CENP-A nucleosome, inhibiting intra- and intermolecular motions and propagates these to the chromatin fiber.

We note that CENP-C expression is tightly regulated, despite over-expression of many centromere proteins in human cancers, including CENP-A (183). Taken in the context of our findings, maintaining the correct ratio between CENP-A and CENP-C *in vivo* might be critical for preserving the mechanical features of centromeric chromatin. This regulation of CENP-C levels may be required, in part, due to the changes in elasticity measured *in silico* and correlated chromatin clustering. Investigating whether CENP-A elasticity is a feature arising from its surprising rapid evolution, or whether it is conserved and co-evolved with kinetochore proteins, will shed light on centromeric evolution. Thus, even at the level of its nanoscale components, the centromere serves as an excellent model to study the evolution of epigenetic systems.

## 3.4 Methods

### 3.4.1 All-atom computational modeling

I built three nucleosomal systems for simulation: the CENP-A nucleosome as described previously (184–186). and the CENP-A nucleosome with one and two rat CENP-C motif (CENP-$C^{CM}$) fragment bound from PDB ID: 4X23 (40). The CENP-$C^{CM}$ fragments were docked onto the CENP-A interface using the CE algorithm (126) of PyMOL (The PyMol Molecular Graphics System). All-atoms molecular dynamics (MD) simulations were performed with software suite GROMACS 5.0.4 (148). The force field employed to model nucleosomes was amber99SB*-ILDN (187) for proteins, amber99SB parmbsc0 (142, 188) for DNA, ions08 (144) for ions, and the TIP3P water model. Unresolved 3AN2 residues Thr 79 through Asp 83 of CENP-A',

Chain E, were built with the program MODELLER (145). During energy minimization of this constructed region, one residue in the N-terminus and C-terminus directions were unconstrained. Additionally, selenomethionine residues were altered to methionine through a single atom mutation from Se to S. As a control, the 146 base pair α-satellite DNA (147) was aligned onto 3AN2 using the CE algorithm (21) of PyMOL (148). Systems (2) and (3) were built by docking the CENP-$C^{CM}$ fragment from the recently solved structure of an H3 chimera nucleosome bound to CENP-$C^{CM}$ onto the final 1 microsecond snapshot of simulation (1) which was then subsequently run for an additional microsecond.

From these initial structures, the GROMACS tool pdb2gmx was used to assign charges to residues at biological pH. Then, a rectangular cuboid box was created such that boundaries were a minimum distance of 1.5 nm from the unsolvated system. Next, $Na^+$ and $Cl^-$ ions were introduced to neutralize the system charge and model an ionic physiological concentration to 150 mM NaCl. For both preproduction and production runs, periodic boundary conditions were employed. Electrostatics were handled with the Particle Mesh Ewald method and Verlet cutoff scheme. For the non-bonded interaction shift functions, Coulombic and van der Waals potentials had a cutoff distance at 1.0 nm. The covalent bonds to hydrogen were constrained with the LINCS algorithm.

Each system was energy minimized using steepest descent to a maximum energy of 100 kJ/mol. The systems were then equilibrated in multiple steps. First, the

systems were heated to 300 K for 2000 ps. During this step, DNA was restrained with K = 1000 kJ $mol^{-1}$ $nm^{-2}$ in the Canonical ensemble (NVT). For the next thermal equilibration at 300K for 2000 ps, both DNA and protein had weak harmonic position restraints K = 2.5e-5 kJ $mol^{-1}$ $nm^{-2}$ to prohibit large nucleosome rotations. Lastly, pressure was equilibrated for 1500 ps in the Isothermal-isobaric, NPT, ensemble at 300 K and 1.0 bar.

Production simulations were performed for 1 microsecond at 300 Kelvin. Temperatures were V-rescaled which is velocity rescaling with a stochastic term to achieve the proper ensemble (150) with a time constant of 1.0 ps. System pressures were regulated with the Parrinello-Rahman barostat (189) at 1.0 bar and a time constant of 2.0 ps. The simulations' time step size was 2 femtoseconds. Coordinates, velocities, and energies were saved every 2 ps. Non-bonded neighbors lists were updated every 20 femtoseconds. For subsequent analysis, trajectories were truncated to remove the first 600 nanoseconds to account for additional system equilibration during production runs. I performed structural analysis calculations, PCA, contact analysis, RMSF, as described previously (152).

For quality control and checks for equilibration, the energy minimization, equilibration, and running RMSD for the simulations (*Appendix B*, Fig. B4). Both CENP-A and CENP-A with one and two CENP-$C^{CM}$ bound (40) ran for an additional microsecond and the first 600 ns of simulation time were truncated from the dataset for further analysis and to account for equilibration. For a control to compare to this

dataset, I also analyzed the H3 nucleosome from previous work (126). In addition to our prior description, after energy minimization I checked structures for potential clashes based on van der Waals radii through the accepted range of 0.4 – 1.0 Å and verified that there were no clashes in the nucleosome structures.

### 3.4.2 Computational calculations of gaping and sliding

Furthermore, I calculated the relative positions of three phosphate backbone atoms at positions -33, -43, and +38 numbered from the 5' (−) to 3' (+) direction relative to the pseudo-dyad as previously marked in FRET experiments to measure gaping and sliding (23). The distances between these points and the skew of the triangle formed were measured and then plotted with the initial position of residue -33 set to (0,0) on a xy-plane. The distribution of $\Delta y$ and $\Delta x$ of +38 relative to -33 and -34 was used to measure DNA gaping and sliding respectively. I visualized these distributions with standard box plots showing the mean, the interquartile range, and whiskers extending to the extrema. The distribution of polygons contains the minima and maxima of all three vertices were plotted visually with triangles to present changes in skew and the range of sizes. Comparative shifts in DNA motions towards gaping and sliding were used to show a trend towards those motions, but with lesser magnitude of motion compared to experiments, since the simulation timescale is orders of magnitude smaller than the corresponding experimental time scales.

### 3.4.3 *In silico* calculation of Young's modulus

Minimal Cylinder Analysis (MCA) will be briefly introduced here, as applied to the data presented in Chapter 3. For a detailed form, MCA is presented in Chapter 4. MCA was further refined between the work covered in the two chapters—in Chapter 3 there were three input variables. In Chapter 4 the model was refined to include only two. The goal of this analysis is to model each nucleosome as a homogenous elastic "minimal" cylinder for each time step of the simulation, retrieve the cylinder height and radius distributions, and from this data calculate the *in silico* Young's Modulus of the nucleosomes. My method to calculate the dimensions of the minimal cylinders follows the workflow:

[1] Orient the nucleosomes so that they lie "flat" on the x-y plane. To achieve this, I calculated the principal axes of the moment of inertia, where the first principal axis defines the broadest plane of the nucleosome. The axes of symmetry of the nucleosomes align with the three principal axes, $p_1$, $p_2$, $p_3$, with the center-of-mass at the origin.

[2] Calculate the surfaces of the cylinder so that they coincide with stiffer regions of the nucleosomes. I addressed this issue by calculating the root mean square fluctuations (RMSF) of each residue along the simulation since the structural disorder of a region positively correlates with local structural fluctuations. Since RMSF is a time-averaged parameter, multiple timesteps are required to calculate fluctuations of

residues. As a result, I divided the simulation into windows (800 windows per simulation) and calculated the RMSF for each residue in each window.

[3] Retrieve the average heights, radii, and the variances of these distributions. To do so, I sorted the C-α coordinates by their z-axis coordinates and selected the z coordinate of the residue where ten stiffer residues below an RMSF threshold were excluded outside of the cylinder volume. From the height, h, and radius, r, data I calculated the average h and r, the variance or spread of the distributions, and the standard deviations Δr and Δh.

[4] The outputs from step [3] then served as the variable inputs to calculate the Young's Modulus of each system. The work done in the deformation of an elastic material is stored in the form of strain energy, which I calculate for the deformation of the cylinder in the absence of the shear stresses. In my simulations, the amplitude of vibrations depends on the amount of energy given to the system from the temperature, or the thermal bath of the solvent. From equipartition theorem, $1/2\ k_bT$ (where $k_b$ is the Boltzmann constant and T is temperature, 300 K) is the amount of energy attributed to the observed cylinder deformation. From the data on the average cylinder conformation, the magnitude of elastic deformation, and the energy input from the thermal bath I calculate the Young's modulus. I calculated the standard deviation of Young's modulus values from three independent subsections of the analyzed trajectories. This method is presented in detail in Chapter 4 with further refinement.

# Chapter 4: Minimal Cylinder Analysis reveals the mechanical properties of oncogenic nucleosomes

*This chapter is adapted with modification from: Mary Pitman, Garegin A. Papoian, Yamini Dalal. "Minimal Cylinder Analysis Reveals the Mechanical Properties of Oncogenic Nucleosomes". Biophys J. 118(9):2309-2318. (2020).* (167)

Author Contributions: Conceptualization of MCA: MP, GAP; Conceptualization of Oncogenic Project: YD, GAP. Derivation: MP; Methodology and code: MP; Investigation: MP; Writing: MP, GAP, and YD.; Funding Acquisition: GAP, and YD; Visualization: MP; Supervision: GAP and YD.

## 4.1 Introduction

The elastic properties of chromatin regulates genetic function in a manner distinct from classically understood properties such as key binding partners (13). Early evidence of the elastic behavior of chromatin comes from the classic micromanipulation experiments of grasshopper chromosomes (190). Several subsequent studies have shown that chromatin acts as an elastic medium and that its constituent linker DNA behaves as an entropically driven elastomer (191). Such studies on the physics of chromatin have led to new biological insight. For example, the pericentromere, a region flanking the site of microtubule attachment, can act as a mechanical spring, governing chromosome separation and spindle length during mitosis (192, 193). Chromatin physics can help to address questions about chromatin

ordering (194), how DNA is both stable and distortable (34, 156, 195), how glassy DNA dynamics give rise to cell-to-cell variability (136, 196), and even how the mechanical micro-environment tunes genetic expression (197). Chromatin states are altered by posttranslational modifications (PTMs) (11, 198) and by histone variant deposition at the macromolecular scale (40). Consequently, the additive effects of nanoscale modifications are an essential component of chromatin chemical signaling pathways and may alter chromatin's mechanical behavior.

In Chapter 3, I investigated the material properties of CENP-A nucleosomes and binding partners, located primarily at centromeres, and H3 nucleosomes found throughout the chromosome arms (136). More specifically, to gain new insights into the initial effects of kinetochore formation, I determined the Young's moduli of centromeric nucleosomes and the tuning effects of binding partners that seed kinetochore formation (9). This report presents my new algorithm to perform elasticity measurements *in silico*, denoted Minimal Cylinder Analysis (MCA), and outlines the theory and derivation for how to obtain Young's modulus from stain fluctuations of nucleoprotein complexes.

Furthermore, in this work I have applied MCA to investigate the material properties of hybrid nucleosomes containing simultaneously CENP-A and H3.3 histones. Such nucleosomes are found in human cancer cells and appear to be detrimental to chromosome integrity (9). In many aggressive forms of cancer, CENP-A, a centromeric histone H3 variant, is overexpressed (21, 30). Studies have

demonstrated that either in cancer cells derived from patient tumors, or when artificially over-expressed, excess CENP-A is deposited outside the centromere, and stably retained there in the form of unexpectedly stable (29, 30, 81) hybrid nucleosomes containing CENP-A and H3.3 (21). This complex has been technically challenging to study experimentally due to its low abundance *in vivo*, therefore, motivating us to rely on MCA to explore the material properties and biological impacts of hybrid CENP-A:H3.3 nucleosomes in cancer cells.

Elasticity, as defined by Hooke's Law, is the ability of a material to return to its initial state after deformation by an applied force. The reversibility of this process implies that mechanical energy is stored as elastic strain energy during deformation and is conserved during recoil (29, 30, 199). Additionally, the proportion of stress to strain in the linear regime is described by Young's modulus. Although Young's modulus is a salient mechanical property of a material, it is applicable for small deformations, and, hence, may not be sufficient to predict all biologically germane deformation processes. The function of elastic materials also depends on extensibility and the amount of work required to fracture the material, referred to as toughness (200). For example, exceptionally tough biological materials exist, such as viscid spider silk, which is far more elastic (0.002 GPa) than Kevlar (130 GPa) and yet is remarkably tougher than Kevlar (150 vs. 50) MJ $m^{-3}$ respectively (200).

There exist several computational approaches to model the elastic properties of macromolecules. One such method is finite element analysis (FEA), where a mesh

network describes the structure, and energy is minimized in response to deformation (200). However, the accuracy of this method requires system-specific parameterization to account for atomic interactions such as Coulombic forces. FEA at the nanoscale has produced results consistent with Molecular Dynamics (MD) when informed by atomistic simulation (201, 202), but FEA lacks the built-in portability and resolution of MD. To achieve all-atom resolution, force-probe MD simulations have been implemented (203). However, large systems such as macromolecular complexes are computationally costly, and unphysical force loading rates are typically required due to short simulation time. Lastly, coarse-grained MD force-fields have also been developed that are, excitedly, able to study the non-elastic deformation and fracture of macromolecules to simulate nanoindentation (204, 205). The longer timescales achieved by coarse-grained methods are promising, but they lack the resolution of all-atom and may not resolve differences due to PTMs or variants. In the methodology I present here, I analyze all-atom resolution simulations of nucleosomes at extended time scales, and then use surface fluctuations to derive the modulus of elasticity in the absence of applied forces. The strength of my methodology is that it does not require expensive computational resources beyond equilibrated simulations.

The elastic modulus is derived by connecting equilibrium strain fluctuations with stress response (206). I employ this logic to obtain the elasticity of nucleosomes without applying an external force. Furthermore, I have introduced a simple temporal hierarchy when implementing this algorithm: first, the equilibrium trajectory is

averaged over short timescale windows, and the resulting structures snuggly fit into encompassing cylindrical bounding domains. Afterward, the sequence of fluctuating cylinders is analyzed using solid mechanics, while also estimating the energy of the corresponding low-frequency vibrational mode from the equipartition theorem. Overall, this algorithm produces the absolute values of nucleosomes' Young's moduli, without freely adjustable parameters that are tuned to fit experiments.

## 4.2 Methods

The goal of this analysis method is to calculate the Young's modulus of nucleosomes in the absence of applied forces. Essentially, this technique connects structural fluctuations observed in unbiased MD simulations, with the nucleosome's mechanical response. To analyze all-atom simulation data in such a way, I first treat the nucleosomes as mechanically homogenous elastic cylinders vibrating in a thermal bath. Next, I calculate the dimensions and fluctuations of what we term "minimal" cylinders over the ensemble of each trajectory. I define the cylinder dimensions as the minimum volume that contains the rigid exterior surfaces of the nucleosome.

To develop a simplified model for elasticity calculations, I make assumptions based on the known physical properties of nucleosomes. First, I apply an averaging technique to all-atom simulation data using continuum mechanics. By doing so, the nucleosome is treated as an elastic cylinder. Elastic continuum theory has been shown to predict material properties on the nanoscale when compared to experiments and analytical predictions (207). I further reduce degrees of freedom and variability by

utilizing the pseudo-symmetry and geometry of nucleosomes to treat them as homogenous circular cylinders. Next, I make simplifications on the mode of deformation studied. To compare to single-molecule nanoindentation, I assume that nucleosomes are compressed perpendicular to the axis of the cylinder. Therefore, I model nucleosome fluctuations as compression and expansion in the absence of shearing motions and attribute to this mode an equipartition of energy.

The workflow I used to determine the Young's modulus of nucleosomes from atomistic trajectories is as follows:

4.2.1 Define the all-atom nucleosome coordinate system

4.2.2 Probe for rigid external cylinder bases and lateral surfaces

4.2.3 Retrieve average cylinder dimensions and variances

The output of these steps is then used to calculate the Young's modulus. Next, I will describe in detail each of these steps.

### 4.2.1 Define the all-atom nucleosome coordinate system

Analogous to the requirement of consistent orientation of nucleosomes in nanoindentation studies, I must first choose a standard nucleosome orientation. The question asked is this: if nucleosomes were to lie 'flat' on a surface, what would this orientation be? The chosen alignment influences the cylinder dimensions that will be measured. Since I constrain the analysis to right circular cylinders, if the nucleosomes are tilted, the measured dimensions will be altered. Prior to analysis, I removed the nucleosome's rotational and translational motions from the trajectories and moved the

center of mass (COM) of residues analyzed in MCA to the origin. To define the coordinate system, I computed the principal axes of rotation and oriented the cylinder base to the plane of the first two principal axes (Fig. 4.1a). The axes of symmetry of the nucleosomes align to the three principal axes. I confirmed that the axes were orthogonal and performed a rotational transformation of the atomic coordinates so that the principal axes aligned to the Cartesian coordinate system with the center-of-mass at the origin.

The nucleosome core particles (NCP), which I analyzed with MCA, were built as described in section 2.2 without flexible histone tails. After simulation, during MCA I did not include in the dataset the first and last ten DNA base pairs from simulation so that the NCP fits more snuggly into a cylinder. The coordinates of the protein C-$\alpha$ and nucleic phosphorus atoms were used to calculate the nucleosome dimensions, elaborated next.

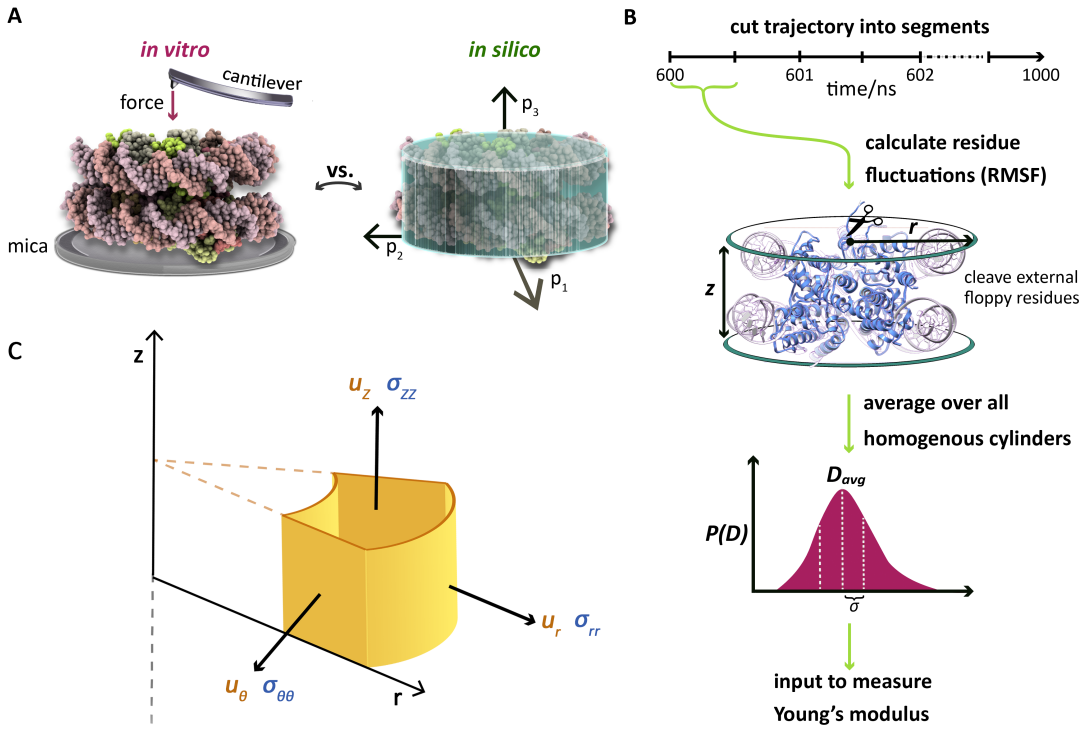


**Figure 4.1: Minimal Cylinder Analysis versus nanoindentation. (a)** Schematic that compares *in vitro* AFM single-molecule nanoindentation force spectroscopy, left, to my *in silico* modeling and analysis, right. In AFM, the applied force from the cantilever is normal to the mica surface. For the computational analysis, nucleosomes were oriented by the principal axes of the moment of inertia and then modeled as homogenous elastic cylinders. **(b)** The workflow for calculating Young's modulus *in silico*. Residue RMSF is calculated for each segment of the simulation to obtain an ensemble of cylinders. The average dimensions, $D_{avg}$, of the radius and height (r, z) and the standard deviation, σ, are then input to calculate Young's modulus. **(c)** A diagram to show the orientation of cylinders in relation to variables introduced in Eq. 4.2 and Eq. 4.12. Displacements ($u_r$, $u_\theta$, $u_z$) are shown in yellow. Stresses in the i-th direction from forces applied in the j-th direction, $\sigma_{ij}$, are shown in blue.

### 4.2.2 Probe for rigid external cylinder bases and lateral surfaces

Since I am measuring elasticity without an applied force, we can consider the following thought experiment: if one were to hypothetically push down on the nucleosome surface, at what point would compression become more significantly

hindered? Such determination would, in turn, suggest the coordinates of the rigid surface of the nucleosome. For example, an intrinsically disordered region or loop is configurationally highly distortable compared with regions comprised of α-helices or β-strands. The elastic moduli of the former structures, which are largely entropic in origin, are expected to be orders of magnitude smaller compared with the latter, hence, are neglected in the subsequent analysis. Thus, I need a specific metric for local stiffness. The rigidity of residues can be quantified by the root mean square fluctuations, RMSF, of each residue throughout the simulation. High RMSF values correspond to increased fluctuation or decreased stiffness. Since RMSF is a time-averaged parameter, multiple time steps are required to calculate fluctuations of residues. Therefore, I divide the 1 μs simulation (from 0.6-1 μs) into a number of temporal segments and output each atom's RMSF per segment.

There are two input parameters in MCA: the 'Averaging Window Length' (AWL), which defines the length of the temporal segments; and the 'Flexible Exterior Residue Number' (FERN), which defines how many residues over an RMSF value, the 'Residue Flexibility Cutoff' (RFC), are excluded from the cylinder volume. I probed for how sensitive MCA is to AWL (Fig. 4.2a) and FERN (Fig. 4.3a, b). A third parameter, the 'Residue Flexibility Cutoff' or RFC value, demarcates rigid and flexible residues and is dependent on AWL (Fig. 4.2b). The parameter, RFC, is calculated by MCA. To illustrate this, I plot how the RFC value compares to the atom RSMFs (Fig. 4.2c). I will next explain in greater detail the selection and analysis of these parameters.

### 4.2.3 Parameter sensitivity

To calculate the Young's modulus, I rationally selected both input parameters 'Averaging Window Length' (AWL) and the 'Flexible Exterior Residue Number' (FERN). The AWL parameter specifies the time length of trajectory divisions for MCA. For each temporal segment I calculate local flexibility through the time-averaged parameter, residue RMSF. If the RMSF is less than the 'Residue Flexibility Cutoff' (RFC), a residue is counted as rigid. The RFC value is calculated in MCA based on the RMSF dataset and is dependent on AWL (Fig. 4.2b), described below. The second input parameter, FERN, defines the number of flexible residues cleaved from the exterior surfaces to determine the minimal cylinder dimensions.

I will first address the parameter sensitivity of the algorithm to AWL and its selection. Since AWL divides the trajectory into collections of frames over which RMSF is time averaged, AWL must be large enough to produce RMSF data with a resolvable spread. If the value of the AWL is too small, there is unclear separation of rigid and flexible residues. Excessively long sampling intervals correspond with a statistically insignificant number of generated cylinders. I assessed AWL values ranging from 10 to 400 frames while holding the other input parameter, FERN, constant (Fig. 4.2a). When I fit the Young's modulus output data by linear regression, I found a slightly negative slope (-0.009) but with a low $R^2$ value of 0.012. The standard deviation of the Young's modulus output when varying AWL from 10 to 400 is ± 0.82 MPa. I selected an AWL value of 20 since it falls on the lower end of this range and therefore more cylinders are calculated. The lack of sensitivity to AWL

over this range indicates that the trajectories are long enough to be insensitive to being partitioned into longer sampling intervals over this range. Shorter trajectories may be sensitive to AWL over the range of tested AWL values and should be tested based on the trajectory length.

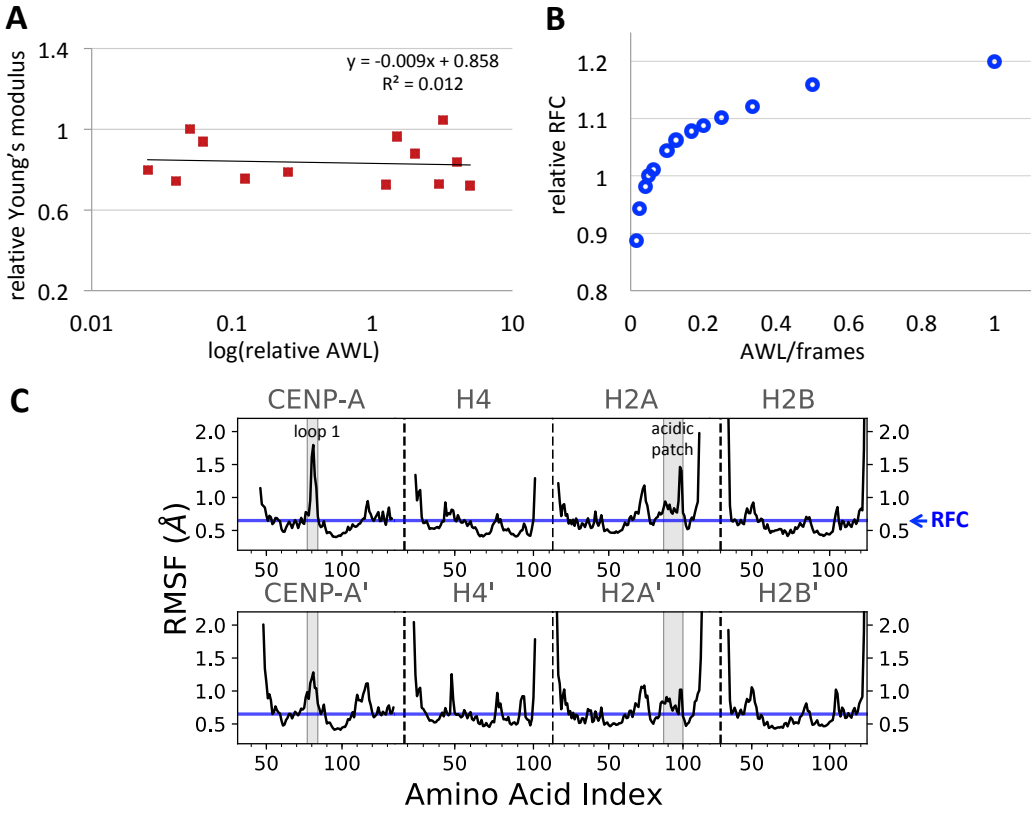


**Figure 4.2: Parameter sensitivity to the Averaging Window Length. (a)** The log of the relative AWL input to the relative output, Young's modulus. The parameter AWL specifies the length of trajectory divisions for MCA. AWL values are restricted to factors of the analyzed trajectory frame number, and so the log is shown due to clustering of factors at low numbers. **(b)** The ratio of the AWL parameter to the number of analyzed frames, where 1 equals the whole trajectory, is shown on the x-axis. The y-axis shows the 50$^{th}$ percentile RMSF value calculated during MCA with varied AWL input. The 'Residue Flexibility Cutoff' (RFC) specifies the maximum RMSF value counted as rigid and is calculated during MCA as the 50$^{th}$ percentile value of the RMSF dataset. **(c)** Example dataset of CENP-A C-α RMSF data at AWL/frames equal to 1 with the RFC value shown as a blue line. Below the blue line residues are counted as rigid.

The RFC specifies the maximum RMSF value counted as rigid and is calculated by MCA as the 50$^{th}$ percentile value of the RMSF dataset. I plotted the relative change in the RFC value based on the number of times the trajectory is divided, the AWL input parameter (Fig. 4.2b). I found a monotonic increase in the RFC value and dependence on the AWL parameter. At point AWL/frames is equal to 1, a singular temporal segment is generated which is equal to the trajectory length and the RFC value is calculated for the whole trajectory (Fig. 4.2b). I show a snapshot of the RMSF distributions for the CENP-A NCP to illustrate how RFC intersects the RMSF data (Fig. 4.2c).

The FERN value was selected so that exterior higher fluctuation residues are cleaved from dimension calculations, and boundary motions are probed versus internal motions. FERN is applied to each cylinder surface. I visualized the resulting cylinder dimensions with varying FERN values (Fig. 4.3a). For parameter sensitivity analysis, I tested a range of FERN values from 1 to 30 (3 to 90 residues for the whole cylinder). For a FERN value of 1-5, I saw a continuous increase in Young's modulus values (Fig. 4.3b, c). I visualized this region, labeled 1, and found that the cylinder boundaries were visibly larger than the nucleosome (Fig. 4.3a). From a FERN value of 5-13 ($0.5 \leq$ relative FERN $\leq 1.3$) I saw a plateau region in Young's modulus output (Fig. 4.3b, c). By visualization of the point labeled 2* in the plateau region, I saw the calculated cylinder appeared to be near the surface of the nucleosome (Fig. 4.3a) Beyond a FERN value of 13, the Young's modulus increases to a maxima of nearly double and then drops (Fig. 4.3b, c). At the top of this peak (Fig. 4.3b) labeled

region 3, I output the calculated cylinder dimensions and found the spike in rigidity was caused by the cylinder boundaries crossing through the center of the DNA double helix (Fig. 4.3a). At the peak, labeled 3, the cylinder was too small, intersected through the middle of the DNA, and the cylinder no longer approximated the boundaries of the nucleosome (Fig. 4.3a). Therefore, to measure the dimensional boundaries of the nucleosome I selected a value of 10 residues, which lies within the plateau region at 2*, where the nucleosome fits more snuggly within the cylinder (Fig. 4.3a). As a control, the parameters FERN and AWL were held constant between NCP systems during analysis.

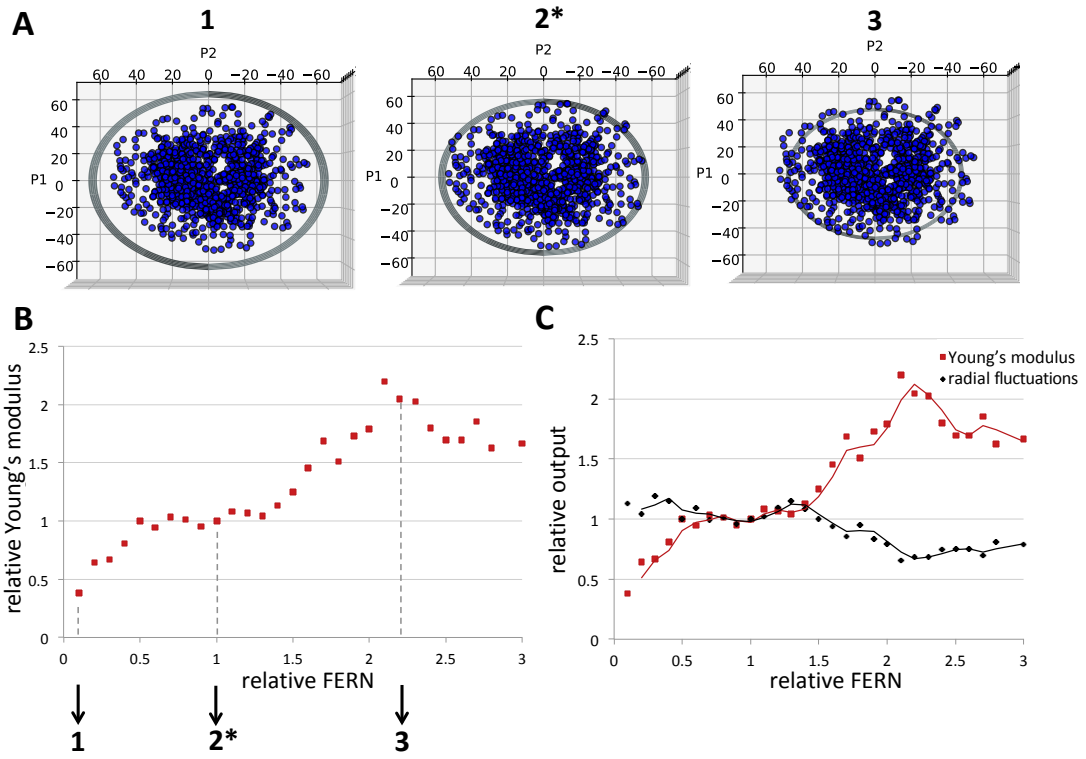


**Figure 4.3: Parameter sensitivity to the Flexible Exterior Residue Number. (a)** The NCP C-α and phosphorus atoms are shown in blue on the plane of the first two principal axes. The cylinder radius is measured from the origin to the grey circular boundary shown. Graph 1 shows the nucleosome and the calculated cylinder dimensions for a FERN value of 1, graph 2* for a FERN value of 10, and graph 3 for a value of 22. **(b)** The relative Young's modulus output dependent on the relative FERN value is shown. The data points graphically shown in panel A are labeled. Point 2* was used for analysis. **(c)** The relative Young's modulus (red) and the relative radial fluctuation, Δr (black), dependent on the relative FERN value.

### 4.2.4 Retrieve average cylinder dimensions and variances

I first sort the C-α and phosphorus atoms by their z-axis coordinates and select the z coordinate of the residue such that the number of residues specified by FERN are minimally excluded outside the cylinder bounds from the top and subsequently the bottom surface. For the analysis of NCPs, I rationally selected a FERN value of 10 residues (Fig. 4.3). Maximal separation among the remaining coordinates

determines the cylinder height, z. I repeat this process for the radial coordinate to calculate the radius, r, of the cylinder per temporal segment. The collection of (r,z) tuples, which are treated as stochastic variables, are histogrammed as illustrated in Fig. 4.3b for an example trajectory. From this stochastic realization for a given simulation I compute the average of the distributions, $z_{avg}$ and $r_{avg}$; and the standard deviations, $\Delta r$ and $\Delta z$. These data are then used in the derivation for the Young's modulus as described below. I have made the MCA code for calculating the Young's modulus available as an open-source tool (208, 209).

### 4.2.5 All-atom computational modeling and analysis

The software suite GROMACS 5.0.4 (210) was used to perform all-atom MD simulations. The force fields used were amber99SB*-ILDN (211) for proteins, amber99SB parmbsc0 (142, 143) for DNA, ions08 (144) for ions, and the TIP3P water model. Two nucleosome systems were built for simulation and compared to prior simulations of H3 as a control (145). First, the heterotypic CENP-A:H3.3 nucleosome was built from the crystallographic structure PDB ID: 3WTP (23). Subsequently, the CENP-A nucleosome was built with PDB ID: 3AN2 (21). The unresolved residues, CENP-A' 79-83, from the crystal structure (3AN2) were inserted using MODELLER (15). Histone tails were not added to the experimentally solved structures. For energy minimization of the inserted residues, the N- and C-terminus were unconstrained. The 146 base pair α-satellite DNA, PDB ID: 3WTP (147), was aligned to all systems, as a control, using the CE algorithm (21) of PyMOL (148).

I used Gromacs tool pdb2gmx to assign residue charges at biological pH around 7 (212): a charge of +1 on Lys (side chain pKa = 10.67) and Arg (side chain pKa = 12.10), 0 for Gln, −1 for Asp (side chain pKa = 3.71) and Glu (side chain pKa = 4.15), and His is neutral and is protonated on the ε-nitrogen (side chain pKa 6.04) (213). The boundaries of the simulation were set to a cuboid box with a minimum distance of 1.5 nm from the nucleosome with periodic boundary conditions. Counter ions of Na+ and Cl- were introduced to neutralize the system and to model an ionic physiological concentration of 150 mM NaCl. The Particle Mesh Ewald method was used for electrostatics with the Verlet cutoff scheme. Coulombic and Van der Waals potentials were used for non-bonded interactions with a cutoff distance at 1.0 nm. The LINCS algorithm was used to constrain covalent bonds to hydrogen.

Energy minimization was performed using steepest descent to a maximum energy of 100 kJ/mol. Following this, equilibration of the structure was carried out. The systems were heated to 300 K for 2000 ps with a DNA position restraint of K = 1000 kJ $mol^{-1}$ $nm^{-2}$ in the Canonical ensemble. Following this, thermal equilibration was performed for both DNA and protein at 300K for 2000 ps with weak position restraints defined as $K_{het}$ = 2.1e-5 kJ $mol^{-1}$ $nm^{-2}$ assigned to the heterotypic nucleosome and $K_{cpa}$ = 2.5e-5 kJ $mol^{-1}$ $nm^{-2}$ for the CENP-A nucleosome. These weak position constraints vary based on atom number in each simulation and restrain nucleosome rotations. Finally, the pressure was equilibrated for 1500 ps in the Isothermal-isobaric ensemble at 300 K, 1.0 bar pressure, and weak position constraint (*Appendix C*, Fig. C1a).

Each production simulation was run for 1 microsecond at 300 Kelvin. Simulation temperature was velocity rescaling with a stochastic term to achieve the proper ensemble (150) with time constant 1.0 ps. The Parrinello-Rahman barostat (189) was used for pressure regulation at 1.0 bar, time constant of 2.0 ps. To investigate the possibility of barostat pressure regulation resonating with the thermal fluctuations of the nucleosomal dimensions, I calculated the natural frequency of the system (152), treating the nucleosome as a homogeneous elastic cylinder, and found two orders of magnitude difference between the faster barostat coordinate rescaling frequency and the slower axisymmetric, acoustic deformation mode frequency considered in MCA.

A simulation time step of 2 femtoseconds was used and coordinates, velocities, and energies saved every 2 ps. The non-bonded neighbor lists were updated at intervals of 20 femtoseconds. In order to analyze equilibrated sections of the production runs, the first 600 nanoseconds were not included in analysis. I checked for convergence of the production runs of the heterotypic nucleosome by tracking the change in RMSD from the initial production run configuration (*Appendix C,* Fig. C1b). Published equilibrium analyses of trajectories CENP-A and CENP-A bound to CENP-C are available in *Appendix B,* Fig. B4 (214, 215). Equilibration analysis of the control system, H3, is available in Fig. S2 of Winogradoff *et al.* (9).

Simulation data was truncated to include the final 400 ns for analysis and then

contact analysis was performed. A cutoff distance of 8 Å was used between histone Cα atoms to compare dimer interface formed between CENP-A:CENP-A vs. CENP-A:H3.3. A value of 1 indicates contact at all frames of the analyzed trajectory and a value of 0 indicates an absence of contact during simulation. Principal component analysis (PCA) was performed on the histone core. The magnitude of motion is multiplied by a factor of 5 in the movies to amplify motions for visual clarity.

DNA was analyzed by residue RMSF with errors calculated over three trajectory segments from 600 ns to 1000 ns of total simulation time. DNA end distances were then calculated from the entry to exit ends of the heterotypic nucleosome and the CENP-A nucleosome, compared to H3 (23).

## 4.3 Results

### 4.3.1 Derivation of Young's modulus from MCA

The work done in the deformation of an elastic material is stored in the form of strain energy, U. The strain energy density, u, the energy stored in small volume elements, can be useful to describe variable strains along a body that sum to the total strain energy:

$$U = \iiint_R u(r, \theta, z) r dr\, d\theta\, dz\,. \quad (4.1)$$

Because the extent of cylinder fluctuations is relatively small, in the range of 1.8 to 9.1% of the average radial or lateral dimension, I rely on linear elasticity and small-deformations' theory. Under these conditions, the strain energy density in cylindrical

coordinates can be calculated for low magnitude stresses from arbitrary directions (23) as

$$u = \frac{1}{2}(\sigma_{rr}\varepsilon_{rr} + \sigma_{\theta\theta}\varepsilon_{\theta\theta} + \sigma_{zz}\varepsilon_{zz}) + (\sigma_{r\theta}\varepsilon_{r\theta} + \sigma_{\theta z}\varepsilon_{\theta z} + \sigma_{zr}\varepsilon_{zr}), \quad (4.2)$$

where $\sigma_{ij}$ is the stress in the i-th direction from force applied in the j-th direction and $\varepsilon_{ij}$ is the strain in the i-j plane (Fig. 4.1c). In Cartesian coordinates, the strain energy density of a volume element under arbitrary stresses can be found at Eq. 8.2.18 of a Solid Mechanics reference text (216) and is given as

$$u = \frac{1}{2}\left(\sigma_{xx}\varepsilon_{xx} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zz}\varepsilon_{zz}\right) + \left(\sigma_{xy}\varepsilon_{xy} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zx}\varepsilon_{zx}\right), \quad (4.3)$$

where $\sigma_{ij}$ is the stress in the i-th direction from force applied in the j-th direction and $\varepsilon_{ij}$ is the strain in the i-j plane. Eq. 4.2 can be derived from and is in the same form as Eq. 4.3. In the absence of internal torques, stress and strain are both second order symmetric tensors (216). This fact can then be used to understand the origin of Eq. 4.3. As a note, the derivation of Eq. 4.2 from Eq. 4.3 can also be done using trigonometric identities or Einstein summation notation. First,

$$\boldsymbol{\sigma\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}. \quad (4.4)$$

I will take the trace of the matrix product, and so the diagonal elements are

$$(\boldsymbol{\sigma\varepsilon})_{11} = \sigma_{xx}\varepsilon_{xx} + \sigma_{xy}\varepsilon_{xy} + \sigma_{zx}\varepsilon_{zx},$$

$$(\boldsymbol{\sigma\varepsilon})_{22} = \sigma_{xy}\varepsilon_{xy} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zy}\varepsilon_{zy},$$

$$(\boldsymbol{\sigma\varepsilon})_{33} = \sigma_{xz}\varepsilon_{xz} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zz}\varepsilon_{zz}. \quad (4.5)$$

Therefore,

$$\mathrm{Tr}(\boldsymbol{\sigma\varepsilon}) = \left(\sigma_{xx}\varepsilon_{xx} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zz}\varepsilon_{zz}\right) + 2\left(\sigma_{xy}\varepsilon_{xy} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zx}\varepsilon_{zx}\right). \quad (4.6)$$

And from Eq. 4.3,

$$u = \frac{1}{2}\mathrm{Tr}(\boldsymbol{\sigma\varepsilon}) \text{ . (4.7)}$$

The form shown in Eq. 4.7 becomes useful because of the cyclic property of trace. I will consider a transformation matrix, **O,** which is any orthonormal basis:

$$\frac{1}{2}\mathrm{Tr}\left(\left(\mathbf{O}\boldsymbol{\sigma}\mathbf{O}^{\mathbf{T}}\right)\left(\mathbf{O}\boldsymbol{\varepsilon}\mathbf{O}^{\mathbf{T}}\right)\right) \text{ . (4.8)}$$

Since **O** is any orthonormal basis, $\mathbf{O}^{\mathbf{T}}\mathbf{O} = \mathbf{1}$, Eq. 4.8 simplifies to

$$\frac{1}{2}\mathrm{Tr}\left(\mathbf{O}^{\mathbf{T}}\mathbf{O}\boldsymbol{\sigma\varepsilon}\right) = \frac{1}{2}\mathrm{Tr}(\boldsymbol{\sigma\varepsilon}) \text{ . (4.9)}$$

This means that Eq. 4.3 can be written in the form Eq. 4.7, the form of Eq. 4.3 will be invariant to any orthonormal basis set. Therefore, Eq. 4.3 in Cartesian coordinates can be written in cylindrical coordinates as shown in Eq. 4.2.

From Eq. 4.2, since the output from the MCA is in terms of strain, I solve for strains from the stress-strain relations of Hooke 's Law, which can be found from a Solid Mechanics reference at Eq. 6.1.8 (217). When solving for on diagonal stresses, the standard result is obtained that

$$\sigma_{rr} = \frac{E}{(1+\nu)(1-2\nu)}\left[(1-\nu)\varepsilon_{rr}+\nu(\varepsilon_{\theta\theta}+\varepsilon_{zz})\right],$$

$$\sigma_{\theta\theta} = \frac{E}{(1+\nu)(1-2\nu)}\left[(1-\nu)\varepsilon_{\theta\theta}+\nu(\varepsilon_{rr}+\varepsilon_{zz})\right],$$

$$\sigma_{zz} = \frac{E}{(1+\nu)(1-2\nu)}\left[(1-\nu)\varepsilon_{zz}+\nu(\varepsilon_{\theta\theta}+\varepsilon_{rr})\right], \text{ (4.10)}$$

where E is Young's modulus and ν is the Poisson ratio (216). I used a value 0.4 used in prior nanoindentation simulations of macromolecules (216). Using the relations found in Eq. 4.10, in the absence of shear stresses and using Hooke's law, the strain energy density in Eq. 4.2 can also be written in the form

$$u = \frac{\nu\mu}{1-2\nu}(\varepsilon_{rr}+\varepsilon_{\theta\theta}+\varepsilon_{zz})^2+\mu(\varepsilon_{rr}^2+\varepsilon_{\theta\theta}^2+\varepsilon_{zz}^2)\ , (4.11)$$

where $\mu$ is the shear modulus and is related to Young's modulus, E, by $\mu = E / 2(1+\nu)$. For displacements $(u_r, u_\theta, u_z)$ in cylindrical coordinates $(r, \theta, z)$ as shown in Fig. 4.1c (218):

$$\varepsilon_{rr} = \frac{\partial u_r}{\partial r}\ , \qquad \varepsilon_{\theta\theta} = \frac{u_r}{r} + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta}\ , \qquad \varepsilon_{zz} = \frac{\partial u_z}{\partial z}\ . (4.12)$$

In the case of a homogeneous axisymmetric cylinder, the center-of-mass at the origin (Fig. 4.1c), $\frac{\partial u_\theta}{\partial \theta} = 0$ and at the walls of the cylinder $\varepsilon_{rr} = \frac{u_r}{r_{avg}}$. For the case of a homogenous cylinder, I will use the notation $\Delta r$ for the displacement in r, $u_r$. Therefore, in this specific case, $\varepsilon_{rr} = \varepsilon_{\theta\theta}$ in Eq. 4.12. For homogenous cylinders (219):

$$\varepsilon_{rr} = \frac{\partial u_r}{\partial r} = \left.\frac{\partial u_r}{\partial r}\right|_{r=r_{avg}} = \frac{\Delta r}{r_{avg}} (4.13)$$

where $\frac{\Delta r}{r_{avg}}$ is input from MCA. I will then use the relation from Eq. 4.12 that in this case $\varepsilon_{rr} = \varepsilon_{\theta\theta}$ and rewrite. Eq. 4.11 in terms of the Young's modulus, the solvable, and strains to obtain

$$u = \frac{\nu E}{2(1-2\nu)(1+\nu)}(2\varepsilon_{rr}+\varepsilon_{zz})^2+\frac{E}{2(1+\nu)}\left(2\varepsilon_{rr}^2+\varepsilon_{zz}^2\right)\ , (4.14)$$

Eq. 4.14 can then be simplified by factoring to arrive at Eq. 4.15:

$$u = \frac{E}{2(1+\nu)}\left[\frac{\nu(\varepsilon_{zz}+2\varepsilon_{rr})^2}{(1-2\nu)}+\varepsilon_{zz}^2+2\varepsilon_{rr}^2\right]. (4.15)$$

Strain values in Eq. 4.15 are calculated from the measured quantities $r_{avg}$, $z_{avg}$, $\Delta r$, and $\Delta z$ using minimal cylinder analysis (Methods 4.2).

I next focus on the acoustic cylindrical mode of motion that describes compression in the z-axis along with radial extension (and *vice versa*). The average potential energy of this mode can be estimated from the equipartition theorem, $U=\frac{1}{2}k_bT$, where $k_b$ is the Boltzmann constant and T is the simulation temperature, 300 K. I then integrate Eq. 4.15 over the body volume, Eq. 4.1, and with the above-mentioned energy from equipartition theorem I solve for Young's modulus:

$$E = \frac{k_bT(1-\nu-2\nu^2)}{V(\varepsilon_{zz}^2-\nu\varepsilon_{zz}^2+2\varepsilon_{rr}^2+4\nu\varepsilon_{zz}\varepsilon_{rr})} . \quad (4.16)$$

### 4.3.2 Experimental validation of MCA model

For the analysis in Chapter 3 (220), the Hertz model with spherical indenter geometry for Young's Modulus measurements was used (9). The Hertz model assumes that the substrate is an isotropic, elastic solid and is valid for small indentations and low forces, in the linear regime. To check for elastic dependence on the point probed, we previously measured the Young's modulus across mononucleosomes and found that the effective elasticity is surprisingly homogenous across the surface (*Appendix B*, Fig. B1) (221). This finding is consistent with the model of MCA, which treats nucleosomes as homogenous elastic solids.

### 4.3.3 Experimental validation for *in silico* Young's modulus calculations

In our study of the mechanical properties of nucleosome variants on the chromatin fiber, in Chapter 3, I applied MCA to measure the Young's modulus of nucleosomes (9). I measured the elasticity three systems: (1) the canonical nucleosome, H3, (2) the centromeric variant of H3, CENP-A, (3) and CENP-A

nucleosomes bound to CENP-C. The *in silico* algorithm to determine Young's modulus quantitatively agrees with *in vitro* nanoindentation measurements (Fig. 4.4a). Thus, experimental nanoindentation studies provide a validation for our model and suggest that the various assumptions made in MCA are acceptable simplifications.

### 4.3.4 Young's modulus of the hybrid CENP-A:H3.3 nucleosome

After validating MCA against *in vitro* single-molecule force studies previously (9), I apply this method to study the elastic properties of a heterotypic cancer-specific nucleosome. This unique variant nucleosome has one copy of CENP-A and one copy of H3.3, and is enriched at CENP-A ectopic sites in chromatin (9), some of which are well documented fragile sites in the chromatin fiber (29). The heterotypic nucleosome was found to be surprisingly stable, regardless of the unique docking interface formed between two divergent H3 variants (30). What, then, causes the measured stability?

To explore this question, I computationally assessed the elastic properties of the heterotypic nucleosome. I discovered an intermediate Young's modulus of hybrid CENP-A:H3.3 nucleosomes (8.5±0.5 MPa) compared to CENP-A nucleosomes (6.2 ±0.4 MPa), and canonical H3 nucleosomes (9.8 ± 0.7 MPa, Fig. 4.4a). Values for nucleosome dimensions and standard deviations are provided for a trial from this dataset (*Appendix C*, Table C1). This result contradicts the idea that unfavorable contacts may form between the CENP-A:H3.3 heterodimer and disrupt the stability of

the hybrid nucleosome. Since this methodology uses an averaging technique over the structure of the nucleosome, I next asked how the dynamics of the heterotypic nucleosome gives rise to its intermediate elasticity. Two hypotheses were considered: first, the heterotypic nucleosome presents an averaged global shift in nucleosome dynamics; or secondly, there may be sequestered regions within the heterotypic nucleosome that display canonical or centromeric nucleosome dynamics.

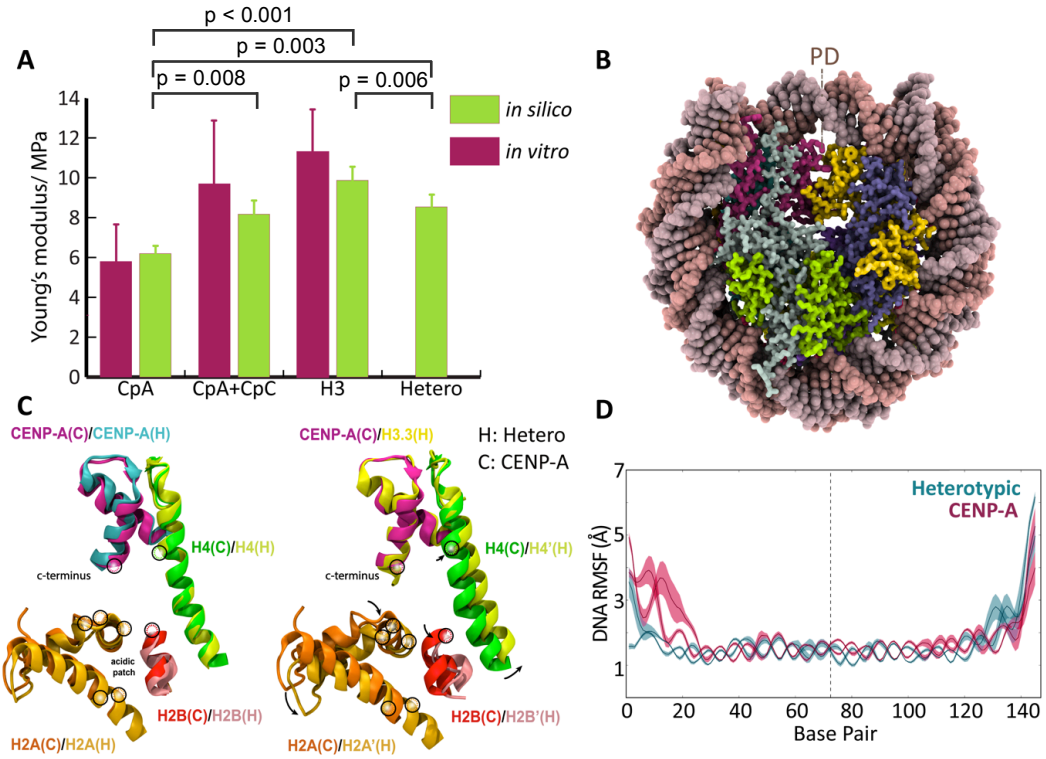


**Figure 4.4: The elasticity of CENP-A histones is preserved in the oncogenic variant.** **(a)** Young's modulus of CENP-A nucleosomes, CpA; CENP-A bound to CENP-C, CpA+CpC; the canonical nucleosome, H3; and the CENP-A:H3.3 containing heterotypic nucleosome, hetero. AFM measurements *in vitro* are shown in magenta and *in silico* measurements are shown in green. Experimental values are referenced from Chapter 3 (21). P-values are calculated for *in silico* measurements for an alpha value of 0.05. **(b)** The structure of the heterotypic nucleosome. Histones CENP-A are shown in magenta, H3.3 in yellow, H4 in dark slate-blue, H2A in light grey, and H2B in green. **(c)** The overlay of the CENP-C binding sites by minimum RMSD obtained from the representative structure of the first principal component (*Appendix C, Supplemental Files,* Movie C1). Histones labeled with a 'C' are from

the CENP-A nucleosome and histones labeled with 'H' are from the heterotypic nucleosome. Black circles indicate CENP-C binding residues. **(d)** Root-mean squared fluctuations (RMSF) of DNA residues of the CENP-A nucleosome, in magenta, and the heterotypic nucleosome, in blue.

### 4.3.5 The rigidified heterotypic nucleosome is permissive to CENP-C binding

The essential docking protein to initiate kinetochore formation is CENP-C, which binds to the surface of CENP-A (9). The structure of the CENP-C binding domain on the surface of the CENP-A nucleosome can be found at Fig. 4.5a. It is still unknown if the heterotypic CENP-A:H3.3 nucleosome is implicated in the formation of neocentromeres. Consistent with reduced elasticity compared to CENP-A (Fig. 4.4a), I found a more tightly bound four-helix bundle interface between H3.3 and CENP-A in the heterotypic nucleosome (*Appendix C*, Fig. C2a). CENP-C docks by interaction with the c-terminal tail of CENP-A in this region and binds across the nucleosome surface with the basic residues of H2A (residues 60, 63, 89-91 in drosophila melanogaster) and H2B (222).

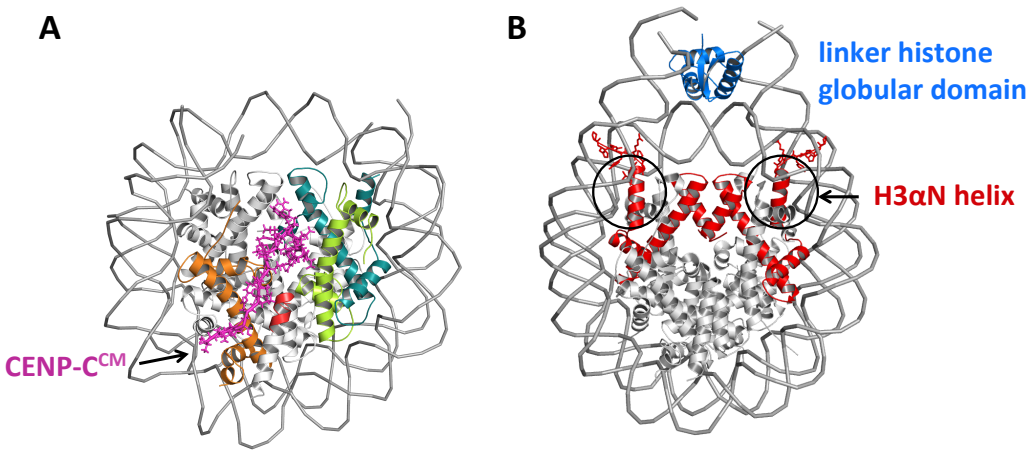


**Figure 4.5: Structural reference of binding partners. (a)** Structural reference of the central motif (CM) binding domain of CENP-C, CENP-C$^{CM}$, bound to the CENP-A NCP. Depicted is the first frame of the CENP-A + CENP-C production run of Chapter 3 (126). CENP-C is shown in magenta, CENP-A in teal, H4 in light green, H2A in orange and the fragment of H2B bound to CENP-C is shown in red. **(b)** Structural reference of the linker histone globular binding domain bound to the canonical nucleosome to form the chromatosome, PDB ID: 4QLC (9). H3 histones are shown in red and the globular domain of the LH is shown in blue.

Therefore, I analyzed the CENP-C binding platform to see if the heterotypic nucleosome is permissive to CENP-C. To do so, I performed Principal Component Analysis (PCA) and animated the first major mode of motion, the first Principal Component (PC1). Visualization of PC1 revealed that the CENP-A containing tetramer of the heterotypic nucleosome rocks apart less than the H3.3 tetramer and is more compact, similar to the CENP-A nucleosome (*Appendix C, Supplemental Files,* Movie C1). Indeed, the heterotypic nucleosome contains asymmetric and partitioned dynamics, where features of CENP-A nucleosome behavior are maintained. In PC1, I see that the CENP-C binding site from the CENP-A nucleosome is preserved in the heterotypic nucleosome (Fig. 4.4c, left). The acidic patch and c-terminal CENP-A domain provide the scaffold for CENP-C binding (223). On the H3.3 face of the

heterotypic nucleosome, the acidic patch, shown as H2A'(H) in light orange, is extended away from the c-terminus of H3.3 (Fig. 4.4c, *Appendix C*, Fig. C2b). This analysis shows that the correct coordination of binding residues for CENP-C is maintained in the CENP-A facing side of the heterotypic nucleosome, making it permissive to the double arginine anchor mechanism of both the CENP-C central domain, R522, R525 (21), previously observed *in vitro* (126), and to the CENP-C motif, R717, R719 (21), which I modeled *in silico* in Chapter 3 (36).

The symmetry breaking in the heterotypic nucleosome across the pseudo-dyad, also propagates away from the CENP-A:H3.3 interface to the RG loop (R80, G81) of CENP-A, L1 (*Appendix C,* Fig. C2). The CENP-A histone displays increased exposure of L1 to solvent in the case of the heterotypic nucleosome in comparison to the CENP-A nucleosome (*Appendix A, Supplemental Files,* Movie C1). For viable kinetochore formation at the heterotypic nucleosome, other proteins such as CENP-N must be recruited (182, 224). The high degree of solvent exposure in the heterotypic nucleosome may indicate that CENP- N is able to bind to its established binding site, CENP-A R80 and G81 (224). It is not yet understood if a single copy of CENP-C and CENP-N are sufficient for kinetochore formation, though minimally, our work on the rigidification of CENP-A when bound to its partners presented in Chapter 3 (9), indicates that these factors would further stabilize CENP-A of heterotypic nucleosomes at ectopic sights.

### 4.3.6 The heterotypic nucleosome pseudo-dyad partitions DNA dynamics

Intriguingly, linker histones (LHs) are excluded from centromere-specific CENP-A nucleosomes, principally because of a clash with entry/exit dynamics of DNA (40). Thus, a fundamental question is whether CENP-A:H3.3 hybrid nucleosomes are able to bind LHs to form a chromatosome unit similar to the canonical nucleosome (225, 226). The LH globular domain docks to the entry-exit sight of canonical nucleosomes (227, 228), illustrated at Fig. 4.5b. The LH disordered tails bind to linker DNA, holding DNA ends together (229). A distinctive difference between H3 nucleosomes and CENP-A nucleosomes is the markedly lower affinity of the latter for LHs (226).

It was experimentally demonstrated that the αN helix in canonical histones binds DNA and restricts DNA end motions, creating a more closed DNA end configuration. In contrast, in CENP-A the αN helix is shorter and experimentally resulted in more open DNA ends and a lack of H1 binding (225). Further, experimentally it was found that unstable entry and exit DNA strands inhibit LH binding (225). A picture of the solved chromatosome structure of the canonical nucleosome bound to the LH globular domain, with the αN helix indicated, is provided for structural reference (Fig. 4.5b) (225). I asked how the intermediate rigidity of the heterotypic nucleosome (Fig. 4.4a) affects DNA dynamics.

First, I calculated the root-mean-square fluctuation (RMSF) of DNA over three segments of our analyzed trajectories. The presence of both CENP-A and H3.3

results in a symmetry breaking in DNA dynamics across the pseudo-dyad (Fig. 4.4d). I observed increased DNA motion in the heterotypic nucleosome proximal to the CENP-A histone in contrast to the H3.3 histone. This region is of interest because the globular domain of H1 binds to the DNA minor groove on-dyad (223).

Furthermore, the asymmetry in DNA dynamics propagates to the DNA entry-exit ends. In the heterotypic nucleosome, I found increased DNA end fluctuations on the end proximal to CENP-A and decreased fluctuations proximal to H3.3 (Fig. 4.4d). The disordered tails of H1 rely on the presence of DNA crossing at the entry and exit ends for nucleosome affinity and to compact the fiber (230). Therefore, I next measured the DNA end-to-end distance in comparison to CENP-A nucleosomes. The DNA ends of the heterotypic are closer together with a probability similar to that of canonical nucleosomes (*Appendix C*, Fig. C2d). More open configurations were least likely to occur by the heterotypic nucleosome (*Appendix C*, Fig. C2d). The increased likelihood of close DNA end configurations suggests that heterotypic nucleosomes may bind LHs.

The somewhat unexpected finding from these simulation analyses is that neither the H3.3 or CENP-A histone dominates the hybrid particle's properties. Indeed, the heterotypic nucleosome displays the dynamics of *both* canonical and centromeric nucleosomes, resulting in an overall intermediate elasticity. Our findings show that the presence of one H3.3 histone variant induces increased rigidity whereas the CENP-A histone intrinsically induces a more elastic phenotype. These results

provide further structural analysis for experimental findings that reported on the surprising thermal stability of the CENP-A:H3.3 containing nucleosome (230, 231).

## 4.4 Discussion

When CENP-A is overexpressed in human cancer cells (21), CENP-A appears to take advantage of H3.3 chaperones, such as HIRA and DAXX (30, 232, 233). The role of H3.3 chaperones in CENP-A deposition away from the centromere provides a logical pathway for the formation of hybrid CENP-A:H3.3 nucleosomes as dimer H3.3/H4 and CENP-A/H4 pairs may fortuitously co-assemble into tetramers on the DNA at regions of high turnover (29, 81). Indeed, H3.3 chaperones are implicated in the ectopic formation of heterotypic nucleosomes in cells with increased survivability in the presence of DNA damage (234–236) and with increased DNAse I sensitivity (29).

The formation and retention of heterotypic nucleosomes on the chromatin fiber could be further augmented by our findings here that CENP-A:H3.3 hybrid is more rigid (8.5±0.5 MPa) than CENP-A alone (6.2 ±0.4 MPa) (Fig. 4.4a). It is important to note that elasticity in the linear regime cannot be extrapolated to either fracture or thermodynamic stability. However, it has been shown that CENP-A nucleosomes are less thermodynamically stable and disassemble more easily than H3 *in vitro* by NAP-1 or heparin destabilization (30, 81), and by magnetic tweezers (231). Intriguingly, the cause for H3 stability was shown to be a more closed DNA

end configuration (237). My results show that the heterotypic nucleosome had the highest likelihood of being in a closed DNA end configuration (*Appendix C*, Fig. C2d). Indeed, the heterotypic nucleosome was measured to have higher thermal stability than CENP-A nucleosomes (231). The heterotypic nucleosome may then be a safe harbor for ectopically located CENP-A histones to be less easily evicted. In other words, the structural features of these hybrid particles might be part of the reason why they persist ectopically, whereas non-hybrid CENP-A nucleosomes may be more readily removed by transcription or remodeling, were they to stochastically accumulate ectopically in normal conditions (21).

Even more fascinating to consider is the dynamics of the heterotypic nucleosome, which is predicted to alter the accessibility of cryptic bindings sites, resulting in downstream biological effects. For example, our findings suggest that heterotypic nucleosomes are competent to bind CENP-C, the structural scaffold for inner kinetochore assembly (238) while still retaining the ability to bind LH H1. The CENP-A:H3.3 containing nucleosome binds the CENP-C central region *in vitro* (182) and ectopic mislocalization of CENP-A results in neocentomeres (21). However, the biological impact of these phenomena depends on the subsequent recruitment of proteins for microtubule attachment, and whether H1 could interfere with the binding affinities or steric space normally available to kinetochore proteins. Minimally, bound LHs or inner kinetochore proteins may further rigidify the heterotypic nucleosome and facilitate CENP-A retention ectopically.

## 4.5 Conclusion

The elasticity of nucleosomes has biological relevance due to the mechanical sensing of large macromolecules and histone variant-specific assemblies such as, in the case of CENP-A, CENP-C and the entire inner kinetochore complex. In the absence of irreversible distortions to the structure, where binding partners or nanomachines exert forces in the elastic range, our newly developed method, MCA, can be applied to measure Young's moduli of various nucleosome complexes that are of low abundance in cells of specific lineages.

Our quantitative elasticity measurements of nucleosomes are likely to also be tunable in varied contexts of biological systems. One process that could alter the elasticity of the nucleosome variants measured here is the presence of binding partners such as proteins, DNA, or RNA. The charge environment, considering the plethora of posttranslational modifications of nucleosomes (30, 81, 199, 233, 239), is also likely to affect the *in vivo* elasticity of nucleosomes and may differ noticeably from the Young's moduli calculated here. Interestingly, nucleosomes are highly responsive to charge perturbations, and slight changes in histone core charge from physiological values thermodynamically destabilize the nucleosome and can cause DNA unwrapping (240). In contrast to these thermodynamic studies, it is still unknown to what extent individual or combinations of site-specific charge modifications alter the elasticity of nucleosomes. Our specific quantitative elasticity measurements are therefore dependent on the environment and state of the nucleosome variant.

The specific structural features of nucleosomes or macromolecules are highly controlled *in silico.* When performing elastic studies *in vitro*, additional sources of potential error arise from structural inhomogeneity and orientation and must be rigorously controlled for (71). In order to make testable predictions *in vivo*, it is then important to consider the specific state of the macromolecule of interest due to increased complexity, inhomogeneity, and cell cycle dependent ensembles of structural states.

A significant benefit of MCA is that it can be applied to equilibrium trajectories, enabling a computationally efficient way to analyze new or existing time-continuous simulations. This has the advantage of providing a means to obtain elastic measurements without doing a series of stretching studies on large systems where single all-atom trajectories are already costly. However, I would like to note that sufficiently long simulations, such the microsecond simulations presented here, are required to produce a sufficient number of uncorrelated segments needed to achieve convergence. In future research, MCA could be extended to different geometries, varied modes of deformation, or augmented to calculate other mechanical properties.

# Chapter 5: Discussion and Outlook

## 5.1 Future directions in the field

We are entering an era coined the Fourth Industrial Revolution (9), typified by automation of manufacturing, smart technology, and increased communication. The transformation to a more technologically advanced society will no doubt also influence the trajectory of scientific research. For example, artificial intelligence (AI) was used to determine optimal synthesis pathways for organic molecules (241) with robotically controlled production. The automation of synthetic chemistry has the power to transform the drug discovery process and outperform human performance. AI's success in organic synthesis begs the question: what processes will be replaced during our scientific careers, and how can we use AI and machine learning (ML) to our advantage?

### 5.1.1 Dimensionality reduction techniques

An outcome of increased communication and ease of information access is an abundance of high-quality biological data. One popular technique that can harness experimental data, maximum entropy modeling, has achieved great success (242), perhaps because the simplest constrained model is deduced. A maximum entropy framework poses a rational and physics-based approach to developing models from experimental data, but it may be computationally costly when fitting to extensive datasets. Existing biological datasets can also train models to predict variables of

interest that are not readily accessible. To predict tissue specific gene expression levels of patients from blood samples, Basu *et al.* trained a linear regression model (243) on Genotype-Tissue Expression data (244). What Basu *et al.* found is that the transcriptome of tissues could be reliably learned from a less invasive procedure—collecting a blood sample. By utilizing large existing databases of experimental data, both theoretical models and real world benefits to patient outcomes can be achieved.

One challenge of dimensionality reduction is interpretability. Interestingly, statistical mechanics can be leveraged through Boltzmann distributions to derive an interpretable model with reduced dimensions. Intensive variables, such as pressure and temperature, are akin to nonlinear dimensionality reduction due to Gibbs-Boltzmann distribution weighting. A constraint-agnostic maximum entropy approach, Thermodynamic Manifold Inference (99, 245, 246), uses the scaffold of statistical mechanics and could offer valuable solutions to solve questions pertinent to genetics. Further, maximum entropy approaches require some artistry when choosing physical variables to constrain.

### 5.1.2 Machine learning to inform simulation

Smart technology is another factor of the 'Fourth Industrial Revolution' (247). Self-Monitoring Analysis and Reporting (SMART) technologies refer to AI, machine learning (ML), and big data analysis so that software can more independently form rational outputs. A subset of this toolbox, ML, will undoubtedly be increasingly implemented to process the vast amount of data at our disposal. Many of us already

integrate concepts of ML into our research. To illustrate this, PCA and MCA are dimensionality reduction methods that help cull from the noise of all-atom simulation parameters of interest.

Repositories for experimental findings pose another big data challenge where a high dimensional space must be filtered to extract parameters. The coarse-grained model AWSEM-MD (241) for example uses the wealth of solved structures available on the Protein Data Bank to provide a structural memory and a biasing potential during simulation. By doing so, the modeled protein essentially learns its fold from existing and known geometries. A statistical method, Direct Contact Analysis (DCA) (74), uses the even greater wealth of protein sequences available to find coevolving amino acid pairs that maintain tertiary structure. DCA inspired AlphaFold (248), which trains neural networks to predict residue distances instead of contact frequency and is prodigious at predicting protein folds.

The chromatin field is also moving towards implementing ML approaches. It was discovered that PTM compartments in the chromatin fiber can be successfully predicted by using ML (249). The genomic compartmentalization of chromatin has also been studied using deep learning. To accomplish this, Ashoor *et al.* translated Hi-C maps to interaction graphs, and the biological functions of lower dimension subcompartments were evaluated by a deep learning classifier (250). Given the immense multivariance of chromatin dynamics in the nucleus, identifying trends in existing data is an area where ML will be of use.

### 5.1.1 Towards a nanomechanical model of chromatin

I determined how specific epigenetic markers such as PTMs and histone variants control the accessibility of binding domains (Chapter 2) (34), alter the free energy landscape (Chapters 2 and 3) (40), and tune the elastic properties of nucleosomes both in isolation (Chapter 3 and 4) (9, 40) and we considered the downstream effects of elasticity in arrays (Chapter 3) (9, 13).

Questions remain about the physics of chromatin and the natural consequences of chromatin's mechanobiology. From my work in Chapter 2, a natural follow-up question arises: how do PTM combinations alter the chromatin fiber? Since histone modifications have unpredictable additive effects (136), it is impossible to extrapolate linearly from disparate PTM modeling data. To fully understand how chemical perturbations alter chromatin dynamics, models sensitive to PTMs that are scalable to arrays will be essential. A potential way forward in this problem would be to map all-atom simulation data, such as the work presented here, onto a coarse-grained model where numerous nonstandard residues, PTMs, are included in the model.

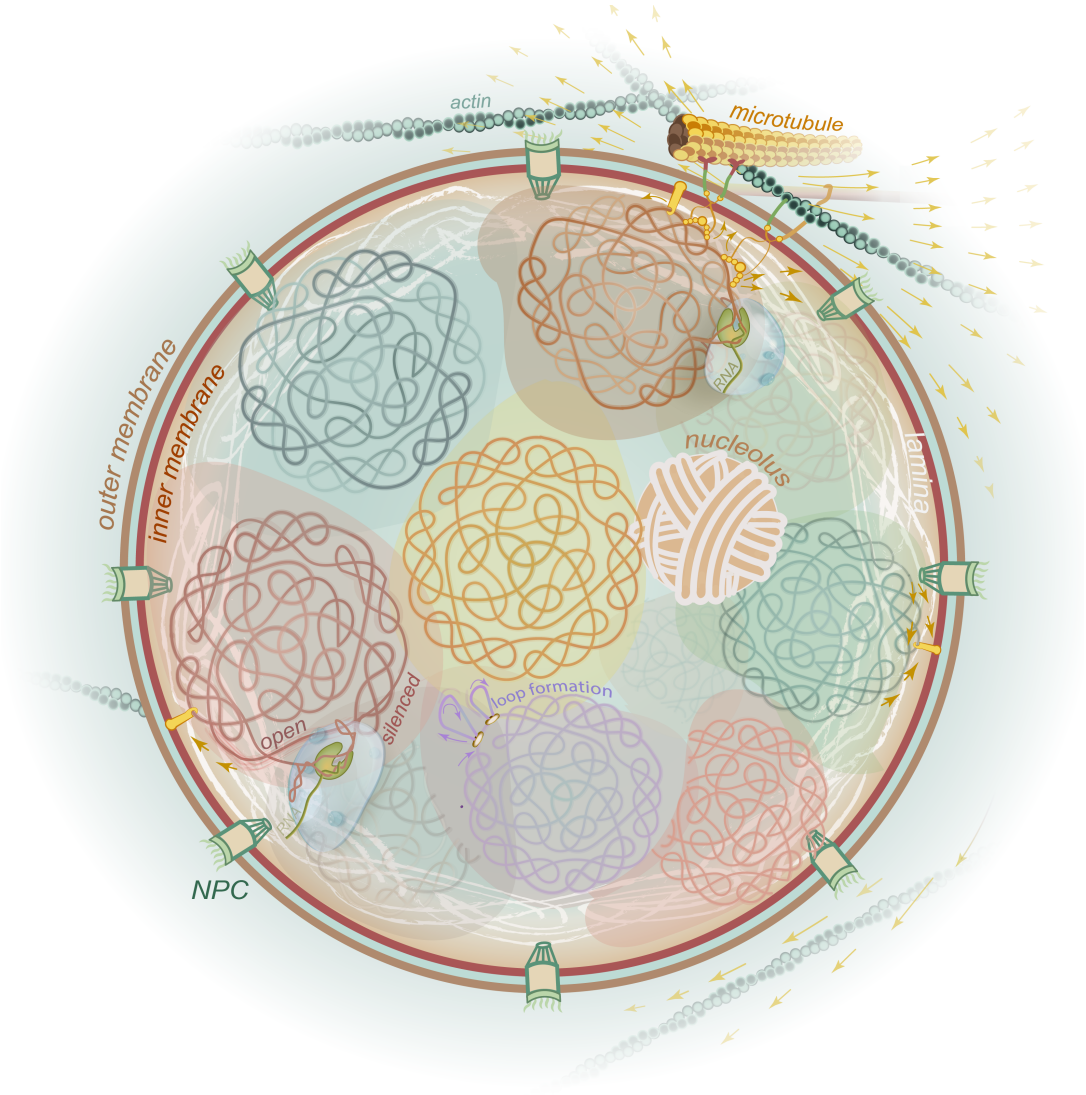


**Figure 5.1: A nanomechanical illustration of the nucleus.** My conceptualization of an interphase eukaryotic nucleus, depicted with chromatin as folded polymers in shaded domains. The coupling of the nuclear membrane and chromatin to external microtubules and actin, labeled, are shown. Vector fields depict active forces. Transcriptional machinery is shown polymerizing RNA within droplets. Loop formation of chromatin is shown with a ring at the base, representing cohesin. Nuclear pore complexes (NPC) are shown embedded in the inner and outer nuclear membrane.

My work presented in the previous chapters infers how nucleosome material properties affect the chromatin array downstream to impact biological function. Broader ranges of stress, strain, and strain rates could be studied to understand how

nucleosomes or chromatin may irreversibly deform during remodeling. Nucleosomes must be evicted for transcriptional machinery to load onto open regions of chromatin, depicted within a liquid-liquid phase separated droplet (Fig. 5.1). Active nonequilibrium processes, such as RNA transcription, thermodynamically stabilize liquid-phase droplets leading to regions of high concentration partially due to viscosity (251). One prediction from the high viscosity and low Reynold's number of nucleoplasm, as discussed in more detail in Chapter 1, is that laminar flow may increase retention of liquid-liquid phase separated droplets during active processes on the chromatin fiber. One such dynamic process in the nucleus is loop formation in chromatin. During interphase, DNA is sequestered into localized regions by loops, mediated by cohesin (Fig. 5.1) and by topologically associating domains (TADs) (252). So, how do epigenetic markers tune the mechanics of chromatin in response these active forces?

Further, when chemical modifications go awry, how is chromatin architecture disrupted, leading to atypical genetic expression profiles? One computational study, published in 2020, found that H3K9me3 modifications lead to denser heterochromatin (253). Chromatin in this condensed state showed preferential positioning towards the nuclear membrane and had nonspecific lamina interactions (253). Lamins are architectural proteins that form within the nuclear membrane (Fig. 5.1) and bind proteins and chromatin (254). This computational study on chromatin positioning provides mechanistic insight into how heterochromatin's transcriptional silencing occurs at the nucleus boundary. But, how location within the nucleus alters

the mechanical properties of chromatin is not fully understood. The nuclear position dependence of gene silencing and chromatin compaction creates one additional layer of complexity for future research.

As computational capabilities continue to grow, we will eventually be able to model chromatin coupled to complex processes such as the interplay between the cytoskeleton and genetic regulation (Fig. 5.1). Ultimately, through the integration of existing physical frameworks and the culmination of our collective efforts, we can revolutionize 'new nuclear physics.'

# Appendix A: Supporting Information for Chapter 2

## A.1 Supplemental Figures

Movies A1 and A2, Movie_A1.mov and Movie_A2.mov, are available under 'Supplemental Files' of the ProQuest portal for dissertation storage.

**Movie A1: Visualization of the first principal mode of histone dynamics is attached in supplemental files.** Histone dimers CENP-A/H4 is shown in red, CENP-A′/H4′ in blue, H2A/H2B in light blue, and H2A′/H2B′ in white. Movies presented here are animations of the most significant mode of motion, $PC1^{core}$, of the principal component analysis of histones ($PCA^{core}$). The first clip shows the histones in both systems rocking with a "freezing" of motion in the acetyl CENP-A histones on the right. The second clip shows the 4-helix bundle in isolation to highlight the interface formed between CENP-A and CENP-A′. Next, the histones are rotated to focus on the described scissoring motion between helices α2 and α3 in the acetyl NCP. Here, the two helices move apart and then together, modulating the widths of the major and minor grooves in the acetyl NCP. Next, flipped to the other side of the nucleosome core, observe the rigidification of the H2A/H2B to H2A′/H2B′ interface in the acetylated nucleosome. To clarify this further, we then show the H2A to H2A′ interface in isolation.

**Movie A2: Visualization of the first principal mode of nucleosome dynamics is attached in supplemental files.** Histone dimers CENP-A/H4 is shown in red, CENP-A′/H4′ in blue, H2A/H2B in light blue, and H2A′/H2B′ in white. Presented are animations of the most significant mode of motion of the whole nucleosome, $PC1^{nuc}$, of the principal component analysis ($PCA^{nuc}$). The pseudo-dyad is labeled PD, and the modified lysine side chains are shown in green with or without acetylation dependent on the system. It is worth noting that our $PCA^{nuc}$ calculations are based on DNA phosphate positions and protein Cα's—therefore, side chains are stagnant relative to the protein backbone. In the first two clips, two unique features of the acetyl NCP are shown: the modulation in the width of the major and minor DNA grooves and the inter-helical DNA bubble formed adjacent to H4 and H4′ K79ac. The final clip then shows the NCP on the side to highlight DNA end untwisting in the acetyl NCP with the last ten base pairs were truncated from the analysis

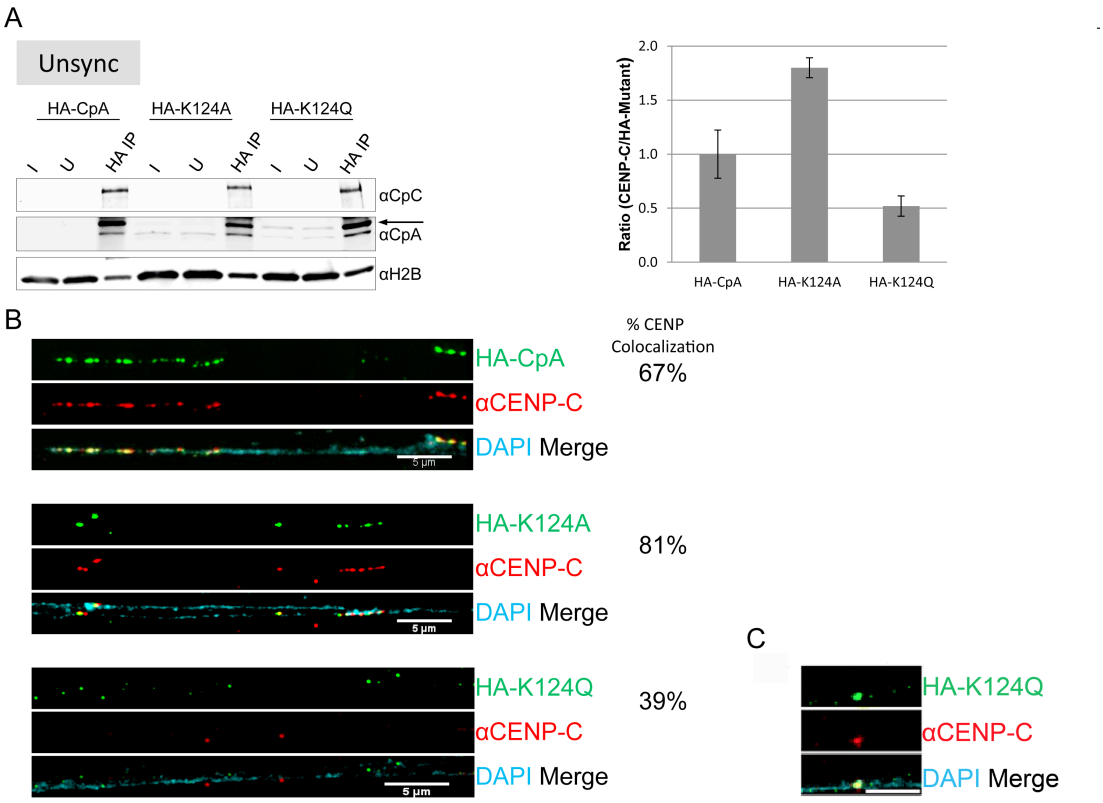


**Figure A1: K124A/Q have altered affinity for CENP-C. (a)** Anti-HA ChIP was performed on medium-sized chromatin arrays (see *Appendix A,* Fig. A.4d), and the amount of co-IP'ed CENP-C (CpC) was determined. *Arrow* indicates location of HA-tagged CpA/K124A/K124Q versus native CENP-A (nCENP-A) below. Three independent replicates conducted using HA-tagged CpA/K124A/K124Q mutants were analyzed by measuring the ratio of CENP-C/ChIP'ed HA-tagged mutant CpA/K124A/K124Q proteins, using LiCor Odyssey linear quantification software. *Bars* indicate standard error of the mean ratio. I = Input and U = Unbound. **(b)** Cells expressing HA-tagged CpA, K124A or K124Q mutants had their chromatin fibers extracted, and immuno-stained against HA (*green*) and CENP-C (*red*) to look for enrichment or depletion of CENP-C on the chromatin fiber (DAPI). Fibers with at least 50% of CENP-C co-localized to the CpA/K124A/K124Q foci were counted and percentage of co-localizing fibers determined. Distribution of the co-localizing fibers is in *Appendix A,* Fig.A1c. **(c)** 'Clumping' K124Q interacts robustly with CENP-C on the chromatin fiber.

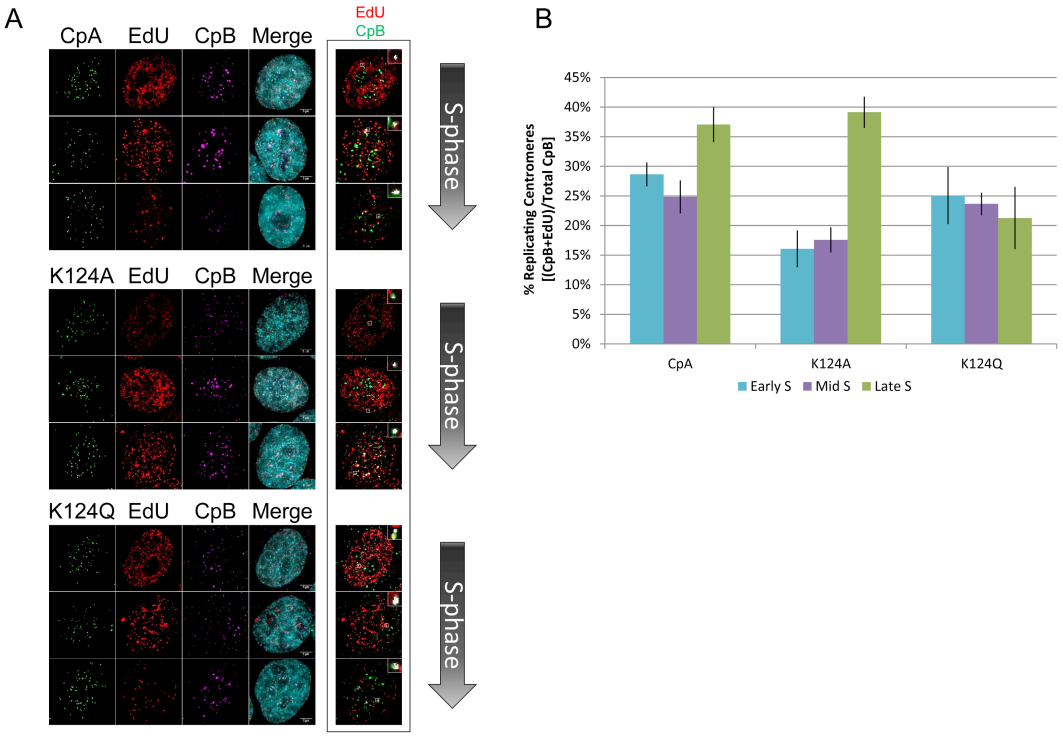


**Figure A2: CENP-A mutations in K124 display altered replication timing of centromeric foci. (a)** EdU-pulsed cells containing CpA/K124A/K124Q were co-stained with CENP-B (CpB) to assess percentage of centromeric replicating origins at early, mid and late S phases. 5 µm *scale bars* are located in the *bottom right* of the merge *panel*. *Boxed* to the *right* is an example of automated co-localization analysis using ImageJ, which was used to determine the fraction of co-localizing CpB and EdU foci (in *white*) with EdU +/CpB + *insets* to show co-localization. **(b)** A graphical representation of the percentage of centromeric origins co-stained with EdU (CpB + EdU) over the total number of centromeric (marked by CpB) foci.

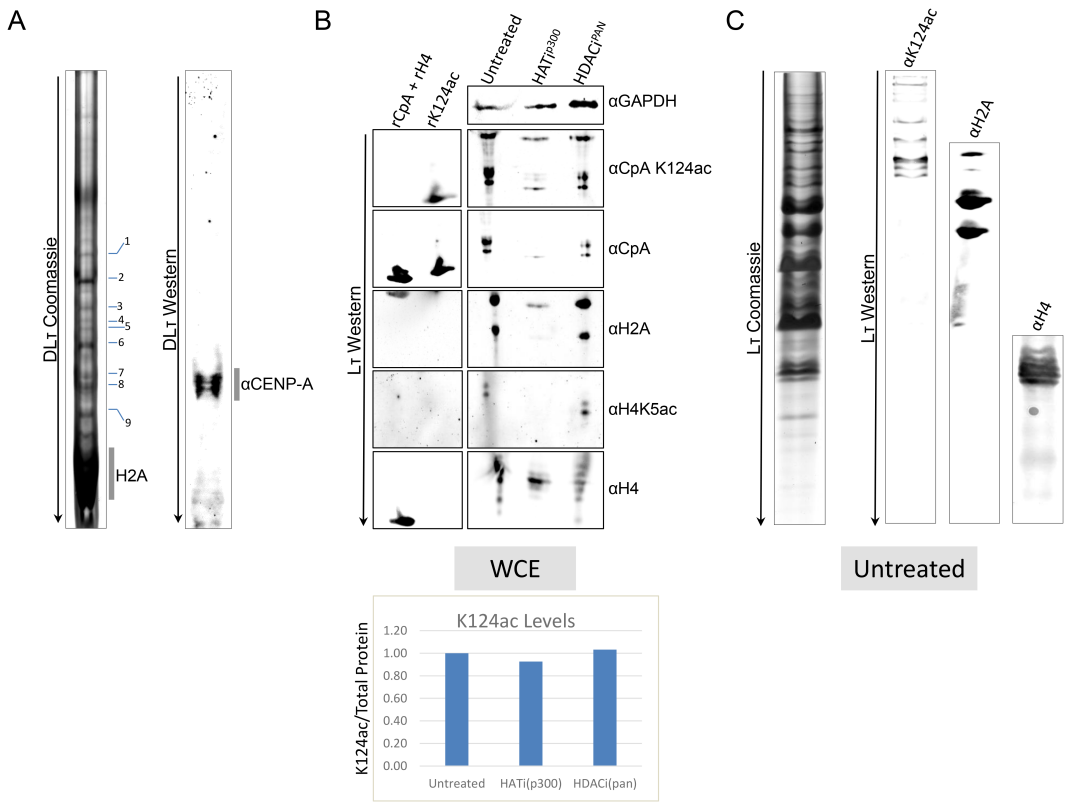


**Figure A3: Identifying CENP-A on double long TAU (DLτ) and long TAU (Lτ) gels for mass spectrometry and HAT/HDAC inhibitor drug treatments. (a)** dLT gels remove excess canonical histone components, leaving behind predominantly CENP-A and histone H2A. Numbers on the DLτ represent bands that were sent for mass spec, and duplicate gel was used for Western and probed against CENP-A. **(b)** Whole cell extracts from cells that were untreated, treated with a HAT inhibitor or HDAC inhibitor, were probed against CpA K124ac, CpA, H2A, H4K5ac and H4. The probes were also used against recombinant CpA/H4 (rCpA + rH4) or recombinant chemically ligated K124ac (rK124ac) to determine antibody specificity. **(c)** Untreated cells with chromatin-bound, hydroxyapatite-purified histones were ran on a long TAU (LT) gel, and duplicate gel transferred to a membrane was probed for K124ac, H2A and H4.

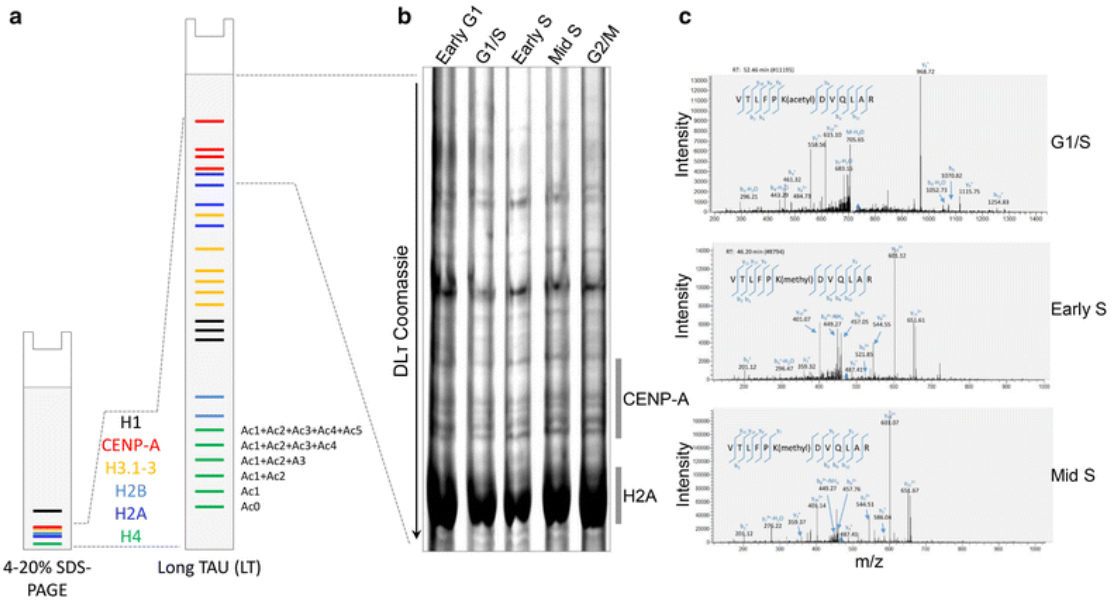


**Figure A4: Endogenous CENP-A K124 is acetylated at G1/S but monomethylated at early S and mid-S phases. (a)** Cartoon depicting the difference between resolving modified histones on a traditional SDS-PAGE versus a long TAU (LT) or double long TAU (DLT) gel. Each additional upward shift of the histone represents an additional acetylated residue. **(b)** G1/S, early S and mid-S chromatin-bound histones were isolated, resolved on a (DLT) gel, stained with Coomassie, and endogenous CENP-A species excised for subsequent analysis by mass spectrometry. **(c)** A peptide containing acetylated lysine 124 was observed in the G1/S sample. The representative MS/MS spectrum showing CENP-A K124 acetylated in the peptide "VTLFPK(acetyl)DVQLAR" is shown on the *bottom left*. Location of the parent peptide ion ($m/z$ = 714.90, charge = +2) prior to fragmentation is indicated in each spectrum with a *blue diamond*. Peptide fragmentation ions identified are indicated in the spectra and on the peptide sequence. The masses of ions b9, b11, y8, y9, and y10 are diagnostic of K124 acetylation. The peptide containing monomethylated lysine 124 was observed in the early S and mid-S phase—the representative MS/MS spectrum showing CENP-A K124 methylated in the peptide "VTLFPK(methyl)DVQLAR" is shown on the *middle* and *top left*. Location of the parent peptide ion ($m/z$ = 466.60, charge = +3) prior to fragmentation is indicated in each spectrum with a *blue diamond*. Peptide fragmentation ions identified are indicated in the spectra and on the peptide sequence. The masses of ions b8, b10, y7, y9 and y10 are diagnostic of K124 methylation

## A.2 Experimental methods

### Cloning

GFP-CENP-A plasmids were a gift from Stephan Diekmann. To mutate K124 to alanine (A) or glutamine (Q) residues, fusion PCR was performed using a reverse primer (TGGGAAGAGAGTAACTCGG) along with a forward primer from the 5′ START codon that includes an EcoRI site. That amplicon was gel purified and combined with a PCR amplicon that used a forward primer encompassing the K to [A] or [Q] mutation (CGAGTTACTCTCTTCCCA[GCG]GATG or CGAGTTACTCTCTTCCCA[CAG]GATG, respectively) and a reverse primer that includes the XmaI site and STOP codon. The final fusion PCR product was excised, gel purified and finally ligated downstream of plasmid that had either GFP or HA-tags, driven by a constitutive CMV promoter.

### Transfections

HeLa cells were grown to ~75% confluency and transfected using Lonza's Amaxa Cell Line Nucleofector Kit R (Cat #VCA-1001) using Amaxa Biosystems Nucleofector II electroporation system according to the manufacturer's guidelines using program O-005. After transfection, cells were plated with fresh media and grown for 48 h before harvesting for experiments.

### Cell synchronization, native chromatin immunoprecipitation (nChIP)

HeLa cells were grown in DMEM (Invitrogen/ThermoFisher Cat #11965) supplemented with 10% FBS and 1X Pen/Strep. nChIP experiments were performed without fixation, and with or without a double thymidine block to synchronize cells. For the complete double thymidine block protocol, please refer to Bui *et al*. (2012).

After cells were harvested, they were washed with PBS and PBS containing 0.1% Tween 20. Nuclei were released with TM2 (20 mM Tris–HCl, pH: 8.0; 2 mM $MgCl_2$) with 0.5% Nonidet P 40 Substitute (Sigma Cat #74385). Afterward, nuclei were washed with TM2 and chromatin was either digested for 4 min for nChIP or 8 min for ChIP-seq with 1.0 U MNase (Sigma Cat #N3755-500UN) in nuclei solubilized with 2 mL of 0.1 M TE (10 mM Tris, 0.2 mM EDTA, 100 mM NaCl) and supplemented with 1.5 mM $CaCl_2$. MNase reactions were quenched with 10 mM EGTA and centrifuged at 1000 rpm at 4 °C. Supernatant was removed, and nuclei extracted overnight at 4 °C in 0.5X PBS supplemented with a protease inhibitor cocktail (Roche Cat #05056489001). ChIP was performed with anti-HA antibody (Santa Cruz Cat #sc-805). nChIP'ed chromatin bound to Protein G Sepharose beads (GE Healthcare Cat #17-0618-02) were washed 3X with cold 0.5X PBS. Westerns were done using LiCor's Odyssey CLx scanner and Image Studio Ver 2.0. Antibodies used for Westerns include: CENP-A (AbCam Cat #ab13939), CENP-B (AbCam Cat #ab25734, CENP-C (MBL Cat #PD030), HA-tag (GenScript Cat #A01244), FLAG-tag (AbCam Cat #ab1162), and H2B (AbCam Cat#52484).

TAU gels and Westerns

For synthesis, preparatory and running conditions for TAU gels, please refer to Walkiewicz, Bui(255). For Western transfers, we used the Trans-blot Turbo Transfer Pack (mini Biorad Cat #170-4158 or midi Biorad Cat #170-4159). For Western detection of the proteins, we used LiCor's secondary infrared antibodies, the Odyssey CLx laser scanning system, and Image Studio Ver 2.0 to quantify the protein levels.

Immunofluorescence (IF)

For complete cell and chromatin fiber IF protocols, please refer to Bui *et al*. (2012)(256). GFP tags were fused in-frame and upstream of wild-type CpA, K124A or K124Q, and may be co-transfected with mCh-CENP-A for co-localization assays. IF was performed using CENP-B (AbCam Cat #ab25734) and CENP-C (MBL Cat #PD030) antibodies. Cells were pulsed with EdU 30 min prior to the desired time point using the Click-iT EdU Alexa Fluor 594 kit (Life Technologies Cat #C10639) and imaged using a DeltaVision RT system fitted with a CoolSnap charged-coupled device camera and mounted on an Olympus IX70. Deconvolved IF images were processed using Image J and to assess co-localization with its 'Colocalization Finder' plug-in.

LC-MS

Tau gel bands were processed using an in-gel digestion protocol from Shevchenko *et al*. (*Nature Protocols* **1**, 2856–2860, 2007). Each band was split in two, and separate trypsin and chymotrypsin in-gel digestions were performed. The samples then underwent shotgun proteomic analysis on a nano-HPLC system

(NanoLC 2D; Eksigent, Dublin, CA) coupled to a hybrid mass spectrometer (Orbitrap Velos Pro; Thermo-Electron, Bremen, Germany). Samples were injected using an auto-sampler and loaded onto a self-packed trap column (2 cm, 100 μm ID, packed with C18 Magic AQ from Michrom Bioresources, Auburn, CA), and the samples were then analyzed on a self-packed C18 (15 cm, 2.7 μm HALO Peptide ES C-18, MAC-MOD, Chadds Ford, PA) column with a laser-pulled tip (P-2000, Sutter, Novato, CA) using a flow rate of 200nL/min. The column was heated to 50 °C using column heater (Phoenix S&T, Chester, PA). Mobile phase A was water with 0.1% formic acid, and mobile phase B was acetonitrile with 0.1% formic acid. The analytical gradient was a 90-min linear gradient from 5 to 35% buffer B. Eluting peptides were electrosprayed at 2.3 kV, and the ion transfer capillary was heated to 250 °C. The Orbitrap was operated in data-dependent mode with different settings depending on the cleavage enzyme used: Trypsin-cleaved samples were analyzed with a CID top 18 method, and chymotrypsin-cleaved samples were analyzed with a CID and ETD decision tree top 12 method. Precursor resolution was set to 60,000, CID collision energy was 35%, and ETD time reaction time was 100 ms with supplemental activation.

# Appendix B: Supporting Information for Chapter 3

## B.1 Supplemental Figures

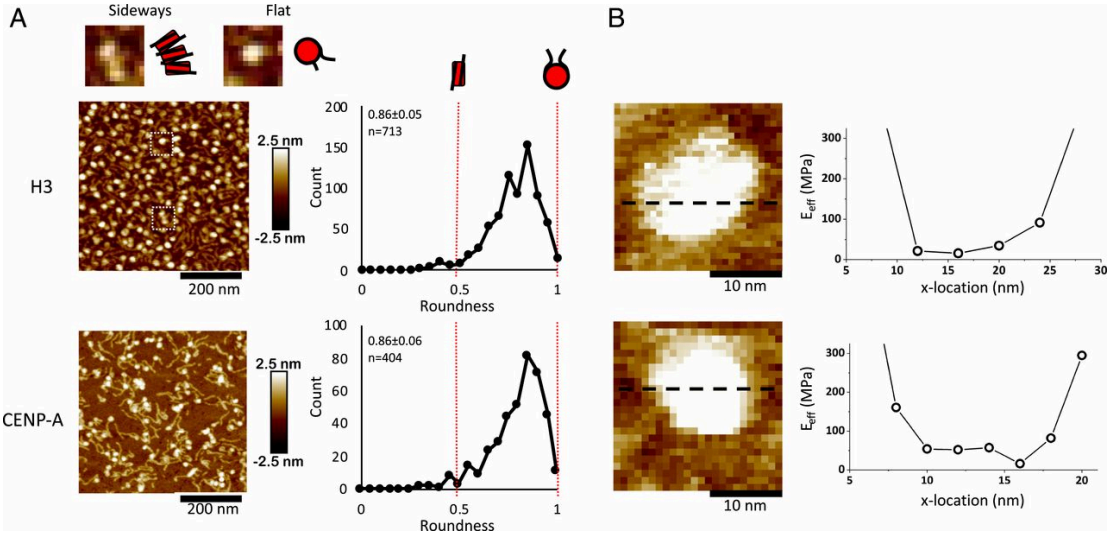


**Figure B1: Mononucleosomes lie flat and are uniformly elastic. (a)** Roundness was measured of either H3 or CENP-A mononucleosomes. A value of 1 would indicate that a particle lies flat on the mica surface and is perfectly round, whereas a value of 0.5 would indicate an oval shape, representing a nucleosome laying on its side. Almost all nucleosomal particles lie flat. Nucleosomes wrap 1.7 turns of DNA, which means that the exit and entry DNA strands are not on the same plane. As a result of this asymmetry, as well as the geometrical limitations of the AFM tip, the nucleosome becomes slightly wedge-shaped with a roundness value of 0.8. (Magnification, 2.5×.) **. (b)** Young's modulus was measured across individual H3 or CENP-A mononucleosomes to assess whether a nucleosome particle is uniformly elastic. No significant difference in Young's moduli was observed across either nucleosome.

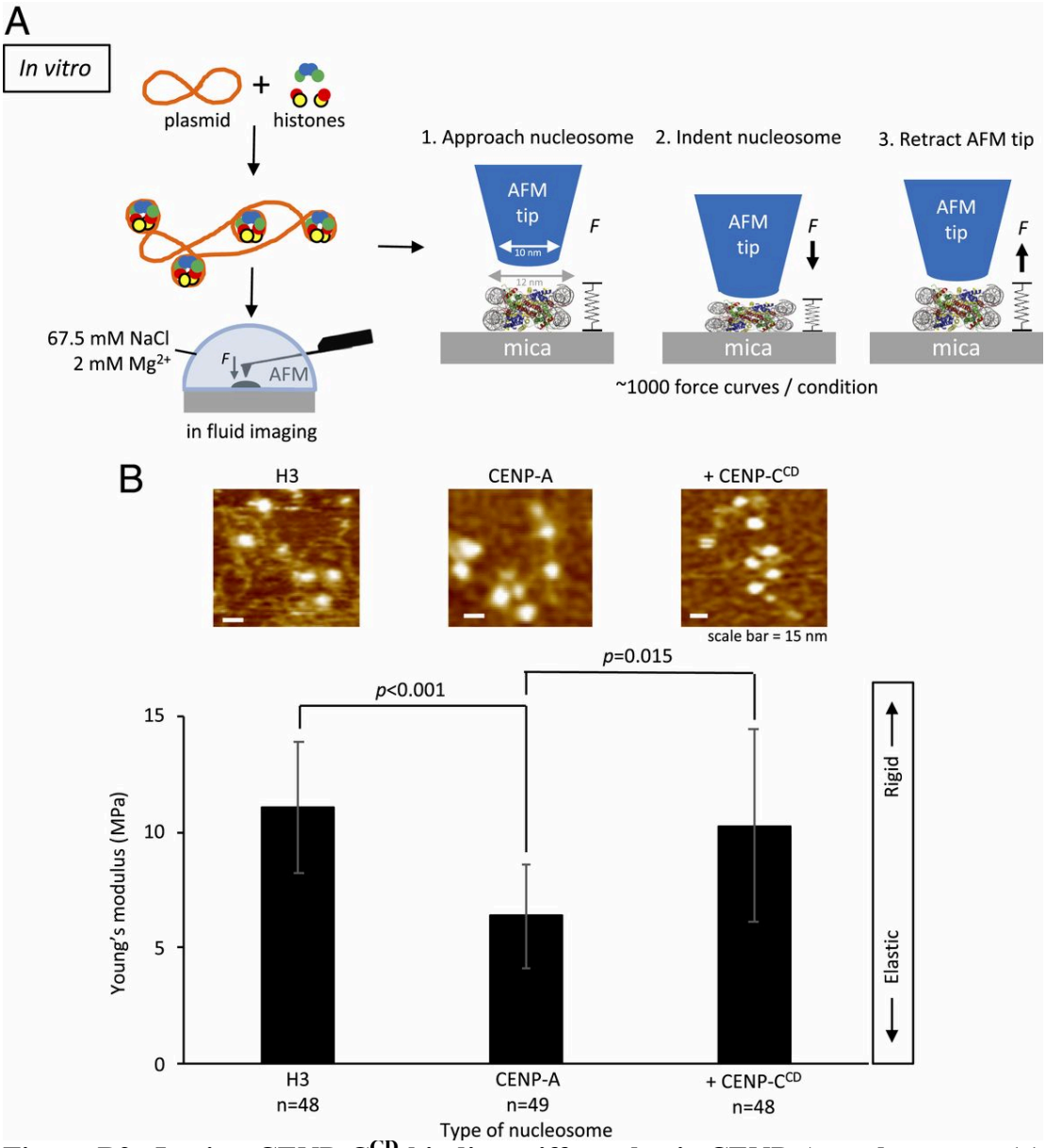


**Figure B2: *In vitro* CENP-C$^{CD}$ binding stiffens elastic CENP-A nucleosomes. (a)** To determine the Young's modulus of CENP-A and H3 nucleosome arrays, we *in vitro*-reconstituted H3 and CENP-A nucleosome arrays by salt dialysis. The AFM tip was aimed at the center of the nucleosome, which were indented by ~1.5 nm under 150 to 200 pN of applied force. Nanoindentation force spectroscopy was performed under near physiological conditions. **(b)** Bar plot summarizing the Young's modulus values showing that CENP-A nucleosomes are more elastic than H3 nucleosomes but become stiffer upon addition of CENP-C$^{CD}$ (2-sided *t* test $P < 0.0001$). Approximately 1,000 force curves were measured per condition.

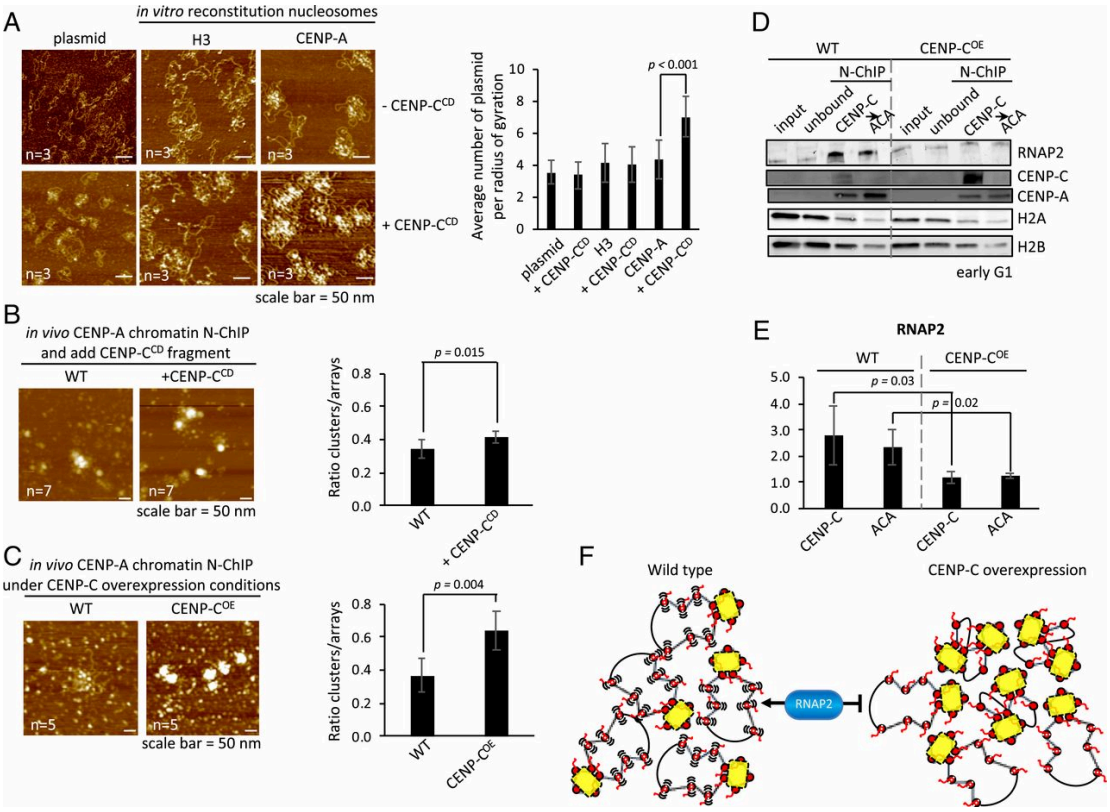


**Figure B3: CENP-C overexpression compacts CENP-A chromatin, making it inaccessible to RNAP2. (a)** Quantitative assessment of *in vitro* reconstituted chromatin showed that only CENP-A chromatin clustered in the presence of CENP-$C^{CD}$ fragment, whereas H3 chromatin or naked plasmids did not in the presence of CENP-$C^{CD}$. Plasmid clustering was measured by counting the number of plasmids in a radius of gyration ($r$ = 0.25 μm). **(b)** To determine if the CENP-$C^{CD}$ fragment used in the *in vitro* experiments could induce CENP-A chromatin compaction, we added CENP-$C^{CD}$ for 30 min to isolated free CENP-A chromatin from HeLa cells. Nucleosome arrays can be identified as either bead-on-a-string or large compacted clusters where DNA strands can be seen entering/exiting. Compacted chromatin was scored over the total number of nucleosome arrays. **(c)** Similar analysis was performed on unbound CENP-A chromatin extracted from cells that either did (CENP-$C^{OE}$) or did not (WT) overexpress CENP-C. **(d)** Centromeres are expressed during early G1. Therefore, we synchronized HeLa cells to early G1 and extracted kinetochore-bound (first CENP-C N-ChIP) and unbound CENP-A chromatin (second ACA N-ChIP of unbound fraction). By Western blot we probed for RNAP2, CENP-C, CENP-A, H2A, and H2B. **(e)** Quantification of RNAP2 levels were determined by LiCor's software. The bar graphs represent 3 independent experiments. **(f)** Working model of CENP-C (yellow) overexpression inducing CENP-A chromatin (red) cross-array clustering thereby reducing access to RNAP2 (blue).

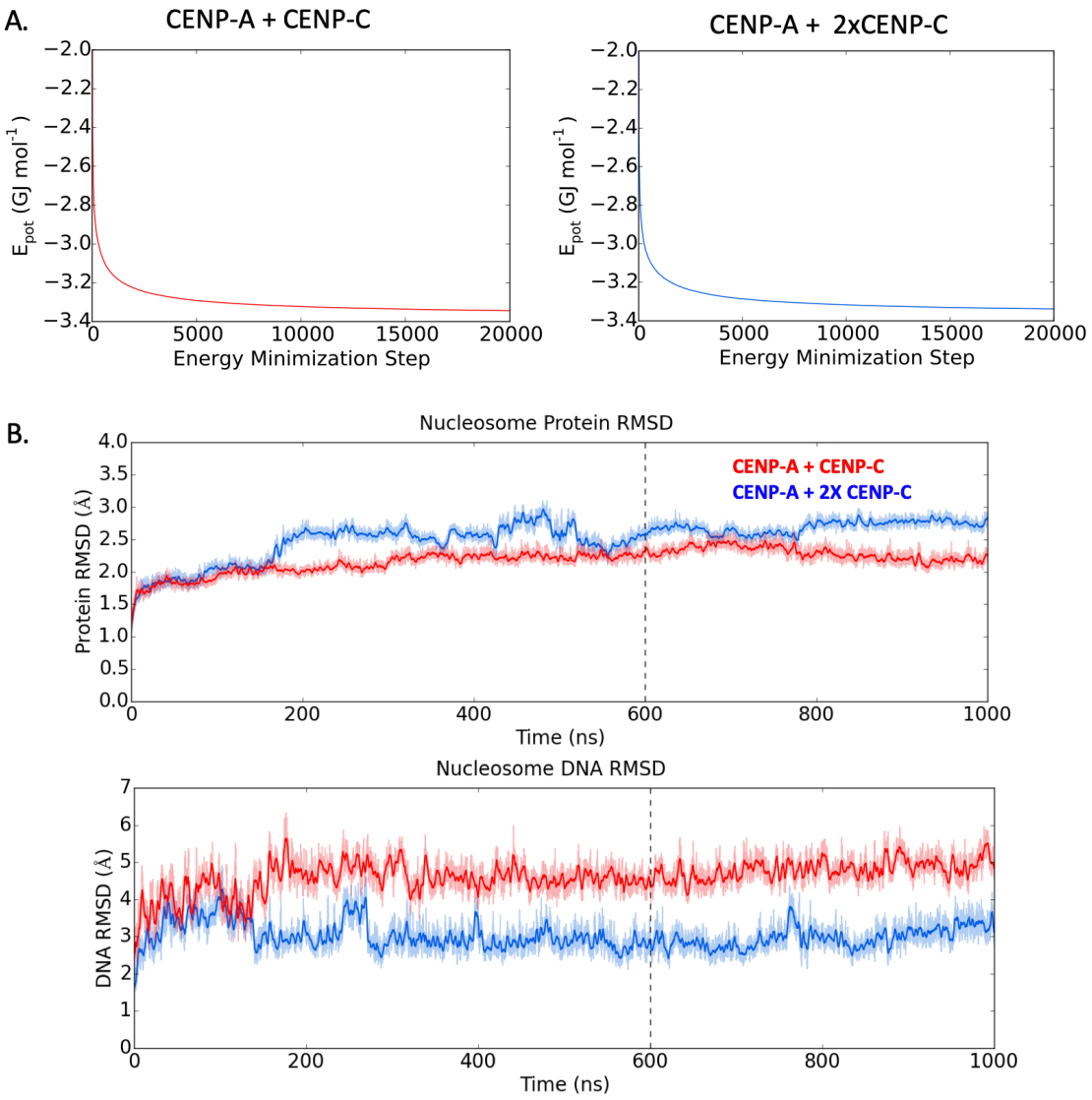


**Figure B7: Quality control of computational simulations**
**(a)** The systems were energy minimized to allow for relaxation of potential clashes or energetically disfavorable rotamers and solvent or ion interactions. **(b)** The simulations ran for 1000 ns and then checked for equilibration by calculation of the root mean square displacement (RMSD) in comparison to the structure after minimization and equilibration. Data before 600 ns was cleaved from the analysis datasets.

**Table B1: Quantification of nucleosomal dimensions by AFM analysis.**
Data demonstrate that *in vitro* chromatin reconstitution yields equivalent dimensions for CENP-A and H3, butthat CENP-A nucleosomes increase in height by ~0.4nm when bound to CENP-CCD. Heights (nm), Diameters (nm),and volumes (nm3) were calculated for representative particles of each class of nucleosome imaged by atomic force microscopy in-fluid conditions (Chapter 3, *Methods*).

**Supplemental Table S1: Quantification of nucleosome particles in H3, CENP-A, and CENP-A + CENP-$C^{CD}$ conditions.**

| H3 | | | CENP-A | | | 2X CENP-$C^{CD}$ | | | 4X CENP-$C^{CD}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| height(nm) | diameter (nm) | volume(nm3) | height(nm) | diameter (nm) | volume(nm3) | height(nm) | diameter (nm) | volume(nm3) | height(nm) | diameter (nm) | volume(nm3) |
| 3.8 | 12.3 | 300.8 | 3.7 | 12.8 | 317.2 | 5.4 | 12.5 | 441.5 | 4.7 | 14.1 | 489.1 |
| 4.1 | 12.3 | 324.6 | 3.7 | 13.4 | 347.6 | 4.5 | 15.1 | 536.9 | 3.6 | 15.1 | 429.5 |
| 3.2 | 15.1 | 381.8 | 3.1 | 12.5 | 253.4 | 4.4 | 12.8 | 377.2 | 5.4 | 12.8 | 463.1 |
| 3.5 | 16.7 | 510.8 | 3.2 | 14.4 | 347.2 | 4.3 | 12.8 | 368.6 | 4.4 | 14.2 | 464.4 |
| 3.7 | 13.5 | 352.8 | 3.2 | 12.6 | 265.8 | 4.3 | 14.1 | 447.3 | 3.7 | 14.8 | 424.1 |
| 3.9 | 14.3 | 417.3 | 3.5 | 13.9 | 353.8 | 4.3 | 12.7 | 362.9 | 3.6 | 12.5 | 294.3 |
| 4.2 | 13.5 | 400.5 | 3.5 | 13.5 | 333.8 | 4.2 | 13.4 | 394.6 | 3.6 | 12.7 | 303.7 |
| 3.9 | 15.5 | 490.3 | 3.6 | 15 | 423.9 | 4.1 | 14.5 | 451.1 | 3.3 | 15.1 | 393.7 |
| 3.9 | 14.3 | 417.3 | 3.6 | 13.8 | 358.7 | 4 | 13.1 | 359.2 | 3.8 | 13.6 | 367.8 |
| 4 | 12.5 | 327.1 | 3.8 | 14.2 | 400.9 | 3.9 | 13.5 | 371.9 | 4.5 | 15.9 | 595.4 |
| 3.7 | 15.5 | 465.2 | 3.9 | 14 | 400.1 | 3.8 | 12.8 | 325.8 | 4.1 | 14.7 | 463.5 |
| 3.7 | 12.7 | 312.3 | 4.1 | 14.1 | 426.5 | 3.8 | 13.6 | 367.8 | 3.8 | 14.1 | 395.6 |
| 3.7 | 13.5 | 352.8 | 4.5 | 13.6 | 435.5 | 4.8 | 11.9 | 355.7 | 3.7 | 17.5 | 593.1 |
| 4 | 12.3 | 316.7 | 4.4 | 12.8 | 377.2 | 4.2 | 13.8 | 418.5 | 3.3 | 15.5 | 414.9 |
| 3.9 | 14.3 | 417.3 | 4.2 | 12.6 | 348.9 | 4.1 | 13.5 | 391.1 | 5.1 | 15.1 | 608.5 |
| 4 | 12.5 | 327.1 | 4 | 12.8 | 342.9 | 3.7 | 14.1 | 384.9 | 4.4 | 16.7 | 642.1 |
| 3.7 | 15.5 | 465.2 | 4 | 14.6 | 446.2 | 3.4 | 12.8 | 291.5 | 4.4 | 12.5 | 359.7 |
| 3.3 | 15.5 | 414.9 | 3.9 | 15 | 459.2 | 3.3 | 15.1 | 393.7 | 4.2 | 13.5 | 400.8 |
| 3.8 | 14.3 | 406.6 | 3.9 | 13.9 | 394.3 | 3.7 | 13.5 | 352.8 | 4.1 | 14.2 | 432.5 |
| 4.1 | 14.3 | 438.7 | 3.8 | 13.8 | 378.7 | 3.9 | 12.9 | 339.6 | 3.9 | 12.6 | 324.1 |
| 3.7 | 15.5 | 465.2 | 3.7 | 14.5 | 407.1 | 4.4 | 12.9 | 383.1 | 3.8 | 13.8 | 378.7 |
| 4.6 | 15.5 | 578.3 | 3.5 | 12.5 | 286.1 | 3.6 | 13.2 | 328.2 | 3.7 | 14.1 | 384.6 |
| 4 | 13.1 | 359.2 | 3.4 | 12.8 | 291.5 | 4.1 | 13.6 | 396.8 | 5.6 | 11.9 | 415.1 |
| 4.4 | 12.7 | 371.3 | 3.3 | 14.2 | 348.2 | 4.2 | 13.9 | 424.6 | 5.4 | 14.3 | 577.8 |
| 3.4 | 13.9 | 343.7 | 4.1 | 13.6 | 396.8 | 4.2 | 15 | 494.5 | 4.4 | 13.9 | 444.8 |
| 3.6 | 12.7 | 303.8 | 4.1 | 14.5 | 451.1 | 4.3 | 13.7 | 422.3 | 4.3 | 13.1 | 386.1 |
| 4.1 | 13.9 | 414.5 | 4 | 12.5 | 327.1 | 4.6 | 14.9 | 534.4 | 4.2 | 11.9 | 311.2 |
| 4.1 | 13.1 | 368.2 | 3.8 | 11.9 | 281.6 | 4.7 | 12.3 | 372.1 | 4.1 | 13.1 | 368.2 |
| 4.3 | 13.9 | 434.7 | 3.8 | 14.9 | 441.5 | 4.9 | 14.1 | 509.8 | 3.9 | 13.9 | 394.3 |
| 3.8 | 14.3 | 406.6 | 3.8 | 13.2 | 346.5 | 4.5 | 12.7 | 379.8 | 5.4 | 13.2 | 492.4 |
| 3.6 | 11.9 | 266.7 | 3.7 | 13.6 | 358.1 | 4.5 | 12.8 | 385.8 | 4.5 | 12.3 | 356.2 |
| 3.8 | 11.9 | 281.6 | 3.7 | 12.8 | 317.2 | 4.5 | 14.2 | 474.8 | 4.1 | 13.2 | 373.8 |
| 3.6 | 13.1 | 323.3 | 3.7 | 12.9 | 322.2 | 4.4 | 13.2 | 401.2 | 4.1 | 14.5 | 451.1 |
| 4.1 | 13.9 | 414.5 | 3.7 | 15.6 | 471.2 | 4.1 | 14.2 | 432.6 | 3.9 | 15 | 459.2 |
| 3.8 | 13.9 | 384.2 | 4.1 | 15.2 | 495.7 | 4.1 | 13.2 | 373.8 | 3.7 | 14.2 | 390.4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4.1 | 15.1 | 489.2 | 3.7 | 12.9 | 322.2 | 3.9 | 14.5 | 429.1 | 3.5 | 13.9 | 353.8 |
| 4.1 | 14.7 | 463.6 | 3.6 | 13.7 | 353.6 | 3.8 | 12.5 | 310.7 | 3.5 | 14.6 | 390.4 |
| 3.5 | 12.1 | 268.1 | 3.6 | 13.6 | 348.4 | 3.7 | 12.9 | 322.2 | 3.4 | 14.1 | 353.7 |
| 3.4 | 13.9 | 343.7 | 3.6 | 16.2 | 494.4 | 3.4 | 14.7 | 384.4 | 3.5 | 13.7 | 343.7 |
| 3.7 | 14.3 | 395.9 | 3.5 | 14.6 | 390.4 | 3.2 | 12.6 | 265.8 | 3.6 | 14.7 | 407.1 |
| 3.4 | 15.5 | 427.4 | 3.4 | 12.7 | 286.9 | 3.6 | 12.7 | 303.8 | 3.8 | 12.9 | 330.9 |
| 4.1 | 13.9 | 414.5 | 3.4 | 15.1 | 405.7 | 3.7 | 14.3 | 395.9 | 3.9 | 14.6 | 435.1 |
| 3.8 | 14.7 | 429.7 | 3.4 | 14.4 | 368.9 | 4.1 | 14.7 | 463.6 | 4.1 | 13.9 | 414.5 |
| 3.7 | 15.9 | 489.5 | | | | 4.3 | 14.2 | 453.7 | 4.2 | 14.1 | 436.9 |
| 3.5 | 14.7 | 395.8 | | | | 4.5 | 12.1 | 344.7 | 4.5 | 13.8 | 448.4 |
| | | | | | | | | | 4.9 | 14.6 | 546.6 |
| | | | | | | | | | 5.9 | 12.7 | 498.1 |

**Table B2: Quantification of chromatin folding demonstrates that CENP-C increases CENP-A chromatin clustering.**

Native chromatin incubated with or without the CENP-C fragment was visually inspected on AFM and identified as "open" or "clustered". Total number of CENP-A clusters/total number of CENP-A nucleosome arrays per scan. Both analyses demonstrate that CENP-C increases CENP-A chromatin clustering (Chapter 3 *Methods*).

| | **ACA WT** | | | **ACA CENP-C$^{CD}$** | | |
|---|---|---|---|---|---|---|
| | # arrays | # clsuters | # clusters/# arrays | # arrays | # clsuters | # clusters/# arrays |
| **Scan1** | 14 | 5 | 0.36 | 19 | 7 | 0.37 |
| **Scan2** | 19 | 6 | 0.32 | 13 | 5 | 0.38 |
| **Scan3** | 22 | 6 | 0.27 | 14 | 6 | 0.43 |
| **Scan4** | 17 | 7 | 0.41 | 26 | 12 | 0.46 |
| **Scan5** | 34 | 13 | 0.38 | 15 | 6 | 0.40 |
| **Scan6** | 48 | 13 | 0.27 | 22 | 9 | 0.41 |
| **Scan7** | 40 | 16 | 0.40 | 24 | 11 | 0.46 |
| | | | | | | |
| | **ACA WT** | | | **ACA CENP-C$^{CD}$** | | |
| | # arrays | # clsuters | # clusters/# arrays | # arrays | # clsuters | # clusters/# arrays |
| **Scan1** | 31 | 10 | 0.32 | 38 | 28 | 0.74 |
| **Scan2** | 34 | 11 | 0.32 | 50 | 27 | 0.54 |
| **Scan3** | 26 | 14 | 0.54 | 49 | 35 | 0.71 |
| **Scan4** | 19 | 7 | 0.37 | 55 | 39 | 0.71 |
| **Scan5** | 28 | 8 | 0.29 | 43 | 21 | 0.49 |

**Table B3: RNAP2 levels on centromeric chromatin are reduced under CENP-C over-expression conditions**

Cells were transfected (or not) with full length CENP-C which was over-expressed (OE) for 3 days,native centromeric chromatin was extracted by CENP-C or ACA ChIP, from wildtype cells or CENP-C OE cells, in parallel. Chromatin was evaluated for RNAP2and CENP-A occupancy on Western blots. 3 independent replicates were quantified using the Licor scanner and automated software. Quantification of RNAP2 in CENP-C IP or ACA IP over Input demonstrates a suppression of RNAP2 levels on centromeric chromatin upon CENP-C OE, and a reduction of total CENP-A levels when RNAP2 is diminished.

**Supplemental Table S3: Quantification of band intensity of RNA polymerase 2 and CENP-A western blots. Values are artibrary units derived from LiCor's software ($10^3$).**

**RNAP2**

| | WT | | | | | CENP-C$^{OE}$ | | |
|---|---|---|---|---|---|---|---|---|
| | Input | CENP-C | ACA | | | Input | CENP-C | ACA |
| Exp1 | 5.92 | 26.8 | 17.5 | | Exp1 | 6.52 | 9.12 | 8.76 |
| Exp2 | 7.76 | 14.1 | 12.6 | | Exp2 | 9.36 | 7.5 | 10.8 |
| Exp3 | 7.34 | 24.9 | 17.8 | | Exp3 | 6.31 | 8.07 | 7.67 |
| | | Ratio CpC/input | Ratio ACA/input | | | | Ratio CpC/input | Ratio ACA/input |
| Exp1 | | 4.53 | 2.96 | | Exp1 | | 1.40 | 1.34 |
| Exp2 | | 1.82 | 1.62 | | Exp2 | | 0.80 | 1.15 |
| Exp3 | | 3.39 | 2.43 | | Exp3 | | 1.28 | 1.22 |
| | mean | 3.25 | 2.33 | | | mean | 1.16 | 1.24 |
| | StDev | 1.36 | 0.67 | | | StDev | 0.32 | 0.10 |
| | | | | | | | | |

**CENP-A**

| | WT | | | | | CENP-C$^{OE}$ | | |
|---|---|---|---|---|---|---|---|---|
| | Input | CENP-C | ACA | | | Input | CENP-C | ACA |
| Exp1 | 0.98 | 8.63 | 13.2 | | Exp1 | 1.97 | 6.74 | 8.53 |
| Exp2 | 0.57 | 10 | 11.4 | | Exp2 | 0.56 | 8.7 | 2.18 |
| Exp3 | 0.38 | 1.71 | 2.23 | | Exp3 | 0.52 | 1.6 | 2.02 |
| | | Ratio CpC/input | Ratio ACA/input | | | | Ratio CpC/input | Ratio ACA/input |
| Exp1 | | 8.81 | 13.47 | | Exp1 | | 3.42 | 4.33 |
| Exp2 | | 17.54 | 20.00 | | Exp2 | | 15.54 | 3.89 |
| Exp3 | | 4.50 | 5.87 | | Exp3 | | 3.08 | 3.88 |
| | mean | 10.28 | 13.11 | | | mean | 7.34 | 4.04 |
| | StDev | 6.65 | 7.07 | | | StDev | 7.10 | 0.25 |

## B.2 Experimental Methods

### Single-molecule nanoindentation of mononucleosomes

H3 (H3 mononucleosome on 187bp of 601 sequence cat#16-2004, EpiCypher, Research Triangle Park, NC) and CENP-A mononucleosome (CENP-A/H4 cat#16-010, H2A/H2B cat#15-0311, 187bp of 601-sequence cat#18-2003, EpiCypher, Research Triangle Park, NC) samples were diluted 1:5 in 2 mM NaCl with 4 mM MgCl (pH7.5) and deposited onto freshly cleaved mica that had previously been treated with aminopropyl-silantrane (APS) as described (135). Samples were incubated on mica for ~3 minutes, excess buffer was rinsed with 400 μL ultrapure, deionized water, and gently dried under an argon stream. Imaging was performed with a commercial AFM (MultiMode-8 AFM, Bruker, Billerica, MA) using silicon-nitride, oxide-sharpened probes (MSNL-E with nominal stiffness of 0.1 nN/nm, Bruker, Billerica, MA). Deposited sample was rehydrated with 10 mM HEPES (pH 7.5), 4 mM $MgCl_2$. Imaging was performed in AFM mode termed "Peak-Force, Quantitative NanoMechanics" or PF-QNM. Images were preprocessed using the instrument image analysis software (Nanoscope v8.15) and gray-scale images were exported to ImageJ analysis software (v1.52i). First nucleosomes were identified as described (79, 257, 258) and subsequently roundness was determined. The Young's modulus was determined by the instrument image analysis software (Nanoscope v8.15).

### Optimization of single-molecule nanoindentation force spectroscopy

Nucleosomes that lay flat, have a round appearance, whereas nucleosomes laying on their side would have an oval appearance. We measured the roundness of both H3 and CENP-A mononucleosomes and found that almost all nucleosomes had a round appearance.

The use of AFM nanoindentation of nucleosomes raise two more concerns. One is that the size of the probe is of the same order of magnitude as the nucleosome. Therefore, widely-used, Hertz-type models used to extract elasticity from indentation data would only provide an effective elasticity that depends on the indentation geometrical parameters such as probe size and precise point of indentation on the nucleosome. This effective elasticity would, however, be comparable between the two types of nucleosomes and their relative values would be comparable to those obtained *in-silico.* The probe sizes used did not vary significantly but we needed to address the possibility that the extracted elasticity depends strongly on the exact point of indentation. If the nucleosome is not uniformly elastic, the precise position of the AFM probe tip could be a critical factor. If the nucleosomes are uniformly elastic, slight differences in where on the nucleosome the elasticity is measured would not be a major concern. We therefore measured the Young's modulus across mononucleosomes. As we are working close to the limit of the instruments noise floor, we considered a Young's modulus variation below an order of magnitude as acceptable. Indeed, measurements of the surrounding mica did show variability greater than an order of magnitude. We found that, in general, effective elasticity did not vary significantly across nucleosomes.

Single-molecule nanoindentation of nucleosome arrays

*In vitro* reconstitution of CENP-A nucleosome arrays (CENP-A/H4 cat#16-010 and H2A/H2B cat#15-0311, EpiCypher, Research Triangle Park, NC) and H3 (H3/H4 cat#16-0008 and H2A/H2B cat#15-0311, EpiCypher Research Triangle Park, NC) on a 3kb plasmid containing a single 601 sequence (pGEM3Z-601 from Addgene #26656) were performed as previously described (79, 257). Human CENP-$C_{482\text{-}527}$ fragment (CENP-$C^{CD}$) (79, 257) and rat CENP-$C_{710\text{-}740}$ (CENP-$C^{CM}$) (ABI Scientific, Sterling, VA) was added in 2.2-fold or 4-fold molar excess to CENP-A nucleosomes. Imaging was performed by using standard AFM equipment (Oxford Instruments, Asylum Research's Cypher S AFM, Santa Barbara, CA). To be able to measure the Young's modulus, the reconstituted chromatin was kept in solution containing 67.5 mM NaCl and 2 mM $Mg^{2+}$ and Olympus micro cantilevers (cat# BL-AC40TS-C2) were used. Before each experiment, the spring constant of each cantilever was calibrated using both GetReal™ Automated Probe Calibration of Cypher S and the thermal noise method (126). Obtained values were in the order of 0.1 N/m. As a reference to obtain the indentation values, the photodiode sensitivity was calibrated by obtaining a force curve of a freshly cleaved mica surface. All experiments were conducted at room temperature. Force-curves for ~50 nucleosomes for all three conditions were measured using both 'Pick a Point' and force-mapping mode. The maximum indentation depth was limited to ~1.5 nm and the maximum applied force was 150-200 pN. For our analyses, we used Hertz model with spherical indenter geometry for Young's Modulus measurements, $\delta = [3(1 - \nu^2)/(4ER^{1/2})]^{2/3}F^{2/3}$ (for a spherical indenter), where $\nu$ is the Poisson ratio of the sample, which is

assumed to be 1/3 as in studies reported previously (259); δ, F, E, and R are the indentation, force, Young's modulus of the sample and radius of the tip respectively. The radius of the tip was confirmed by SEM and found to be about 10 nm in width. Graphs were prepared using ggplot2 package for R.

AFM and cluster analysis

Imaging of CENP-C and CENP-A N-ChIP and bulk chromatin was performed as described (170, 173) with the following modifications. Imaging was acquired by using commercial AFM equipment (Oxford Instruments, Asylum Research's Cypher S AFM, Santa Barbara, CA) with silicon cantilevers (OTESPA or OTESPA-R3 from Olympus with nominal resonances of ~300 kHz, stiffness of ~42 N/m) in noncontact tapping mode or commercial AFM (MultiMode-8 AFM, Bruker, Billerica, MA) using silicon cantilevers (OTESPA or OTESPA-R3 from Olympus). Either *in vitro* samples were exposed to either a) rat or b) human CENP-$C^{CM}$ or c) human CENP-$C^{CD}$ fragments, whereas *in vivo* samples were only exposed to c) hCENP-$C^{CD}$. *In vitro* samples were 1.) naked plasmid DNA, 2.) reconstituted H3 or 3.) reconstituted CENP-A chromatin. *In vivo* samples were kinetochore-depleted chromatin obtained from HeLa cells as described here (79, 257). All samples were incubated for with the CENP-C fragment for 30 minutes at room temperature on an end-over-end rotator, before being deposited on freshly cleaved mica. HeLa cells which transiently transfected with CENP-C were used to isolate kinetochore-depleted chromatin. APS-mica was prepared as previously described (177). The samples were incubated for 10 min, gently rinsed with two times 200 μL ultra-pure water and dried with inert argon

gas before imaging. Plasmid clustering was quantified by counting the total number of plasmids in a 0.25 μm radius of gyration around grouped plasmids. To quantify chromatin compaction, we manually counted chromatin clusters based on their size being at least twice as wide as an individual nucleosome, but with an identifiable entry and exit DNA strand. The cluster were counted over the total number of nucleosome arrays (clustered and not clustered).

Native Chromatin-Immunoprecipitation and western blotting

Human cell line HeLa were grown in DMEM (Invitrogen/ThermoFisher Cat #11965) supplemented with 10% FBS and 1X penicillin and streptomycin cocktail. N-ChIP experiments were performed without fixation. After cells were grown to ~80% confluency, they were harvested as described here (79, 257), but with a few modifications. These were that all centrifugation was done at 800 or 1000 rpm at 4ºC. Chromatin was digested for 6 minutes with 0.25 U/mL MNase (Sigma-Aldrich cat #N3755-500UN) and supplemented with 1.5 mM $CaCl_2$. The first N-ChIP was with 5 μL guinea pig CENP-C antibody, subsequently the unbound fraction was N-ChIP'ed with 5 μL ACA serum (*Appendix B, Methods*). For CENP-C overexpression we transfected HeLa cells with pEGFP-CENP-C using the Amaxa Cell Line Nucleofector Kit R (Lonza cat#VVCA-1001) per manufacturer's instructions. HeLa cells were synchronized to early G1 by double thymidine block (0.5 mM, Sigma-Aldrich cat#T9250). After the first block of 22 hours, cells were released for 12 hours, followed by a second thymidine block of 12 hours. Cells were released for

approximately 11 hours, which corresponds to early G1, based on our previous reports (40, 173, 177).

## Quantification and Statistical Analyses

Significant differences for nucleosome height measurement from AFM analyses and significant differences for immunostaining quantification, and chromatin compaction quantification, were performed using the 2-sided t-test as described in the figure legends and main text. Significant differences for the Young's modulus of *in vitro* reconstituted H3, CENP-A, and CENP-A + CENP-$C^{CD}$ were determined using 1-way ANOVA test using GraphPad Prism software. Significance was determined at $p<0.05$.

# Appendix C: Supporting Information for Chapter 4

## C.1 Supplemental Figures

Movie C1, Movie_C1.mov is available under 'Supplemental Files' of the ProQuest portal for dissertation storage.

**Movie C1: Visualization of the first principal mode of histone dynamics is attached in supplemental files.** Histone dimers CENP-A/H4 are shown in red, CENP-A′/H4′ in blue, H2A/H2B in light blue, and H2A′/H2B′ in white. Movies presented here are animations of the most significant mode of motion, $PC1^{core}$, of the principal component analysis of histones ($PCA^{core}$).

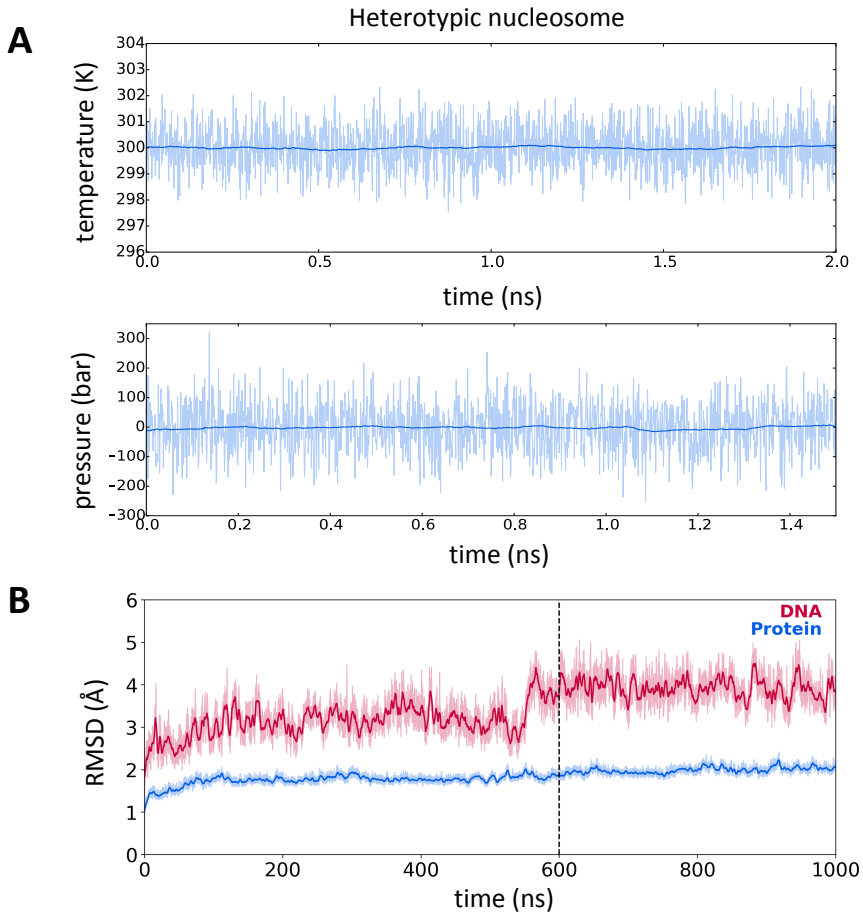


**Figure C1: (a)** The second thermal equilibration to 300 K with weak position restrains, $K_{het}$, of the Heterotypic nucleosome is shown, top. Next, the system pressure was equilibrated to 1 bar, shown on the bottom. **(b)** The RMSD of the Heterotypic DNA and protein compared to the initial production run configuration. We performed analysis from 600 to 1000 ns.

| System | z (Å) | Δz (Å) | r (Å) | Δr (Å) |
|---|---|---|---|---|
| H3 | 55.20 | 0.62 | 54.50 | 0.29 |
| CENP-A | 57.26 | 0.70 | 53.66 | 0.41 |
| +CENP-C | 64.41 | 0.72 | 53.36 | 0.31 |
| Heterotypic | 63.14 | 0.61 | 53.36 | 0.35 |

**Table C1:** Sample trial MCA data for the average dimensions of the height and

radius, $z_{avg}$ and $r_{avg}$, and the standard deviation of the height and radius, $\Delta z$ and $\Delta r$. The elasticity of the structure depends on a nonlinear combination of the ratio of the fluctuation to the dimension size of both r and z.

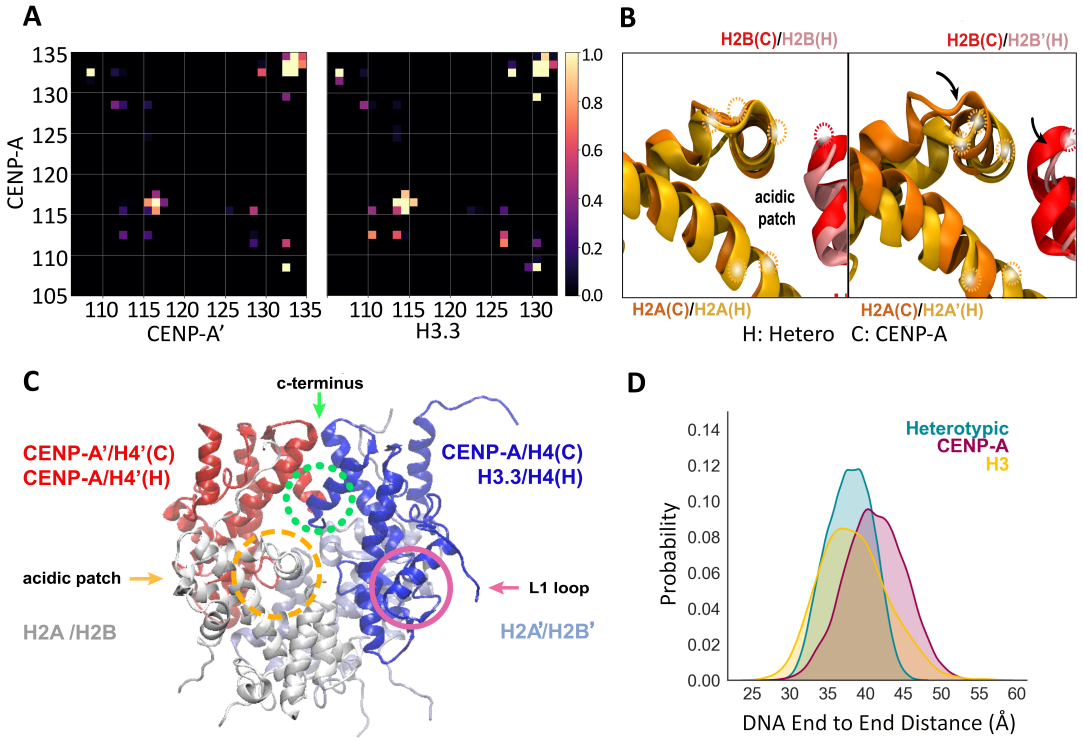


**Figure C2: (a)** Contact analysis at the 4-helix bundle interface of CENP-A:CENP-A in the context of the CENP-A nucleosome on the left in comparison with CENP-A:H3.3 in the heterotypic nucleosome. Increased brightness of color shows the propensity of C-α contact within 8 Å. Black means that contact does not occur, and pale yellow indicates contact at all time-steps. **(b)** The overlay of the acidic patches from the representative structure of the first principal component. System "C" denotes histones from the CENP-A nucleosome and "H" from the CENP-A:H3.3 heterotypic nucleosome. On the left, the CENP-A region from system C is shown with minimum

RMSD alignment to the CENP-A region of the heterotypic nucleosome. For comparison, on the right, the H3.3 domain of H is compared to CENP-A in C. CENP-C binding residues are highlighted. **(c)** Representative image from Movie C1, which depicts histone core motions of the first principal component. Regions of interest in the PCA movie are highlighted such as the histones (where H is the Heterotypic nucleosome and C is the CENP-A nucleosome), the acidic patch in the dashed orange circle, the L1 loop in the pink circle, and the CENP-A or H3.3 c-terminal region in the dotted green circle. **(d)** The histogram of DNA end to end distances for the Heterotypic nucleosome in teal, the CENP-A nucleosome in magenta, the H3 nucleosome in yellow.

# Bibliography


1. S. B. Rothbart, B. D. Strahl, Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* (2014) https:/doi.org/10.1016/j.bbagrm.2014.03.001.
2. K. Luger, M. L. Dechassa, D. J. Tremethick, New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* (2012) https:/doi.org/10.1038/nrm3382.
3. G. D. Bowman, M. G. Poirier, Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem. Rev.* (2015) https:/doi.org/10.1021/cr500350x.
4. J. M. Bellush, I. Whitehouse, DNA replication through a chromatin environment. *Philos. Trans. R. Soc. B Biol. Sci.* (2017) https:/doi.org/10.1098/rstb.2016.0287.
5. N. Petryk, *et al.*, MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science (80-. ).* (2018) https:/doi.org/10.1126/science.aau0294.
6. C. M. Weber, S. Henikoff, Histone variants: Dynamic punctuation in transcription. *Genes Dev.* (2014) https:/doi.org/10.1101/gad.238873.114.
7. P. A. Gómez-García, *et al.*, Mesoscale Modeling and Single-Nucleosome Tracking Reveal Remodeling of Clutch Folding and Dynamics in Stem Cell Differentiation. *Cell Rep.* (2021) https:/doi.org/10.1016/j.celrep.2020.108614.
8. N. Yusufova, *et al.*, Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. *Nature* (2021) https:/doi.org/10.1038/s41586-020-3017-y.
9. D. P. Melters, *et al.*, Intrinsic elasticity of nucleosomes is encoded by histone variants and calibrated by their binding partners. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24066–24074 (2019).
10. G. Nir, *et al.*, Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* (2018) https:/doi.org/10.1371/journal.pgen.1007872.
11. D. Levens, L. Baranello, F. Kouzine, Controlling gene expression by DNA mechanics: emerging insights and challenges. *Biophys. Rev.* **8**, 23–32 (2016).
12. E. de Wit, W. de Laat, A decade of 3C technologies: Insights into nuclear organization. *Genes Dev.* (2012) https:/doi.org/10.1101/gad.179804.111.
13. M. Pitman, Y. Dalal, G. A. Papoian, Minimal Cylinder Analysis Reveals the Mechanical Properties of Oncogenic Nucleosomes. *Biophys. J.* (2020) https:/doi.org/10.1016/j.bpj.2020.01.042.
14. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
15. H. Tachiwana, *et al.*, Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232–235 (2011).
16. O. Perisic, S. Portillo-Ledesma, T. Schlick, Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Res.*

(2019) https:/doi.org/10.1093/nar/gkz234.
17. J. Clausell, N. Happel, T. K. Hale, D. Doenecke, M. Beato, Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS One* **4** (2009).
18. B. Henneman, C. van Emmerik, H. van Ingen, R. T. Dame, Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
19. H. S. Malik, S. Henikoff, Phylogenomics of the nucleosome. *Nat. Struct. Biol* **10**, 882–891 (2003).
20. M. S. Waterman, M. Eggert, A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons. *J. Mol. Biol.* **197**, 723–728 (1987).
21. Y. Arimura, *et al.*, Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci. Rep.* **4** (2014).
22. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *J. Mol. Biol.* **319**, 1097–1113 (2002).
23. D. Winogradoff, H. Zhao, Y. Dalal, G. A. Papoian, Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome. *Sci. Rep.* **5** (2015).
24. X. Huang, W. Miller, A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337–357 (1991).
25. F. Morcos, N. P. Schafer, R. R. Cheng, J. N. Onuchic, P. G. Wolynes, Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc. Natl. Acad. Sci.* **111**, 12408 LP – 12413 (2014).
26. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
27. R. Hirano, *et al.*, Histone variant H2A.B-H2B dimers are spontaneously exchanged with canonical H2A-H2B in the nucleosome. *Commun. Biol.* **4**, 191 (2021).
28. F. Mattiroli, *et al.*, Structure of histone-based chromatin in Archaea. *Science (80-. ).* **357**, 609 LP – 612 (2017).
29. N. Lacoste, *et al.*, Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. *Mol. Cell* **53**, 631–644 (2014).
30. R. K. Athwal, *et al.*, CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics and Chromatin* **8** (2015).
31. S. S. P. Rao, *et al.*, A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
32. E. M. Pugacheva, *et al.*, CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci.* **117**, 2020 LP – 2031 (2020).

33. S. Y. Roth, J. M. Denu, C. D. Allis, Histone Acetyltransferases. *Annu. Rev. Biochem.* **70**, 81–120 (2001).
34. M. Di Pierro, R. R. Cheng, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 12126–12131 (2017).
35. M. Pitman, Y. Dalal, G. A. Papoian, Minimal Cylinder Analysis Reveals the Mechanical Properties of Oncogenic Nucleosomes (2019).
36. M. Milazzo, A. David, G. S. Jung, S. Danti, M. J. Buehler, Molecular origin of viscoelasticity in mineralized collagen fibrils. *Biomater. Sci.* (2021) https:/doi.org/10.1039/D0BM02003F.
37. S. Ling, *et al.*, Combining In Silico Design and Biomimetic Assembly: A New Approach for Developing High-Performance Dynamic Responsive Bio-Nanomaterials. *Adv. Mater.* **30**, 1802306 (2018).
38. K. Popov, J. Komianos, G. A. Papoian, MEDYAN: Mechanochemical Simulations of Contraction and Polarity Alignment in Actomyosin Networks. *PLoS Comput. Biol.* **12** (2016).
39. A. Upadhyaya, Mechanosensing in the immune response. *Semin. Cell Dev. Biol.* **71**, 137–145 (2017).
40. M. Bui, *et al.*, Internal modifications in the CENP-A nucleosome modulate centromeric dynamics. *Epigenetics and Chromatin* **10** (2017).
41. Y. Zhao, B. A. Garcia, Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* (2015) https:/doi.org/10.1101/cshperspect.a025064.
42. K. Karch, J. DeNizio, B. Black, B. Garcia, Identification and interrogation of combinatorial histone modifications . *Front. Genet.* **4**, 264 (2013).
43. J. Williams, *et al.*, MethylationToActivity: a deep-learning framework that reveals promoter activity landscapes from DNA methylomes in individual tumors. *Genome Biol.* **22**, 24 (2021).
44. R. Niwayama, K. Shinohara, A. Kimura, Hydrodynamic property of the cytoplasm is sufficient to mediate cytoplasmic streaming in the Caenorhabiditis elegans embryo. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11900–11905 (2011).
45. D. Boal, *Mechanics of the cell, second edition* (2012) https:/doi.org/10.1017/CBO9781139022217.
46. , Anaphase Movements in the Living Cell: A Study With Phase Contrast and Polarized Light on Chick Tissue Cultures. *J. Exp. Biol.* (1948).
47. J. A. Jacquez, J. J. Biesele, A study of Michel's film on meiosis in Psophus stridulus L. *Exp. Cell Res.* (1954) https:/doi.org/10.1016/0014-4827(54)90144-5.
48. K. S. Bloom, Beyond the code: the mechanical properties of DNA as they relate to mitosis. *Chromosoma* **117**, 103-110doi101007/s00412-007-0138–0 (2008).
49. E. D. Salmon, K. Bloom, Tension sensors reveal how the kinetochore shares its load. *BioEssays* (2017) https:/doi.org/10.1002/bies.201600216.
50. , XXIX. An experimental investigation of the circumstances which determine

whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Philos. Trans. R. Soc. London* **174**, 935–982 (1883).
51. C. M. Caragine, S. C. Haley, A. Zidovska, Surface Fluctuations and Coalescence of Nucleolar Droplets in the Human Cell Nucleus. *Phys. Rev. Lett.* **121**, 148101 (2018).
52. L. Liang, X. Wang, X. Da, T. Chen, W. R. Chen, Noninvasive determination of cell nucleoplasmic viscosity by fluorescence correlation spectroscopy. *J. Biomed. Opt.* **14**, 1–9 (2009).
53. E. M. Purcell, Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).
54. K. Kwapiszewska, *et al.*, Nanoscale Viscosity of Cytoplasm Is Conserved in Human Cell Lines. *J. Phys. Chem. Lett.* (2020) https:/doi.org/10.1021/acs.jpclett.0c01748.
55. T. Qiu, *et al.*, Swimming by reciprocal motion at low Reynolds number. *Nat. Commun.* **5** (2014).
56. P. A. Valberg, D. F. Albertini, Cytoplasmic motions, rheology, and structure probed by a novel magnetic particle method. *J. Cell Biol.* (1985) https:/doi.org/10.1083/jcb.101.1.130.
57. A. R. Bausch, W. Möller, E. Sackmann, Measurement of local viscoelasticity and forces in living cells by magnetic tweezers. *Biophys. J.* (1999) https:/doi.org/10.1016/S0006-3495(99)77225-5.
58. M. T. Valentine, Z. E. Perlman, T. J. Mitchison, D. A. Weitz, Mechanical properties of Xenopus egg cytoplasmic extracts. *Biophys. J.* (2005) https:/doi.org/10.1529/biophysj.104.048025.
59. R. Chambers, H. C. Sands, A dissection of the chromosomes in the pollen mother cells of tradescantia virginica L. *J. Gen. Physiol.* (1923) https:/doi.org/10.1085/jgp.5.6.815.
60. E. G. D'Angelo, SALIVARY GLAND CHROMOSOMES. *Ann. N. Y. Acad. Sci.* (1950) https:/doi.org/10.1111/j.1749-6632.1950.tb39891.x.
61. E. G. D'Angelo, Micrurgical studies on Chironomus salivary gland chromosomes. *Biol. Bull.* (1946) https:/doi.org/10.2307/1538062.
62. S. Na, *et al.*, Rapid signal transduction in living cells is a unique feature of mechanotransduction. *Proc. Natl. Acad. Sci. U. S. A.* (2008) https:/doi.org/10.1073/pnas.0711704105.
63. T. J. Kirby, J. Lammerding, Emerging views of the nucleus as a cellular mechanosensor. *Nat. Cell Biol.* (2018) https:/doi.org/10.1038/s41556-018-0038-y.
64. Y. A. Miroshnikova, M. M. Nava, S. A. Wickström, Emerging roles of mechanical forces in chromatin regulation. *J. Cell Sci.* (2017) https:/doi.org/10.1242/jcs.202192.
65. M. Maurer, J. Lammerding, The Driving Force: Nuclear Mechanotransduction in Cellular Function, Fate, and Disease. *Annu. Rev. Biomed. Eng.* (2019) https:/doi.org/10.1146/annurev-bioeng-060418-052139.
66. G. Xue, *et al.*, Effect of Microenvironment on Differentiation of Human Umbilical Cord Mesenchymal Stem Cells into Hepatocytes *In Vitro* and *In Vivo*. *Biomed Res. Int.* **2016**, 8916534 (2016).

67. B. W. Booth, *et al.*, The mammary microenvironment alters the differentiation repertoire of neural stem cells. *Proc. Natl. Acad. Sci.* **105**, 14891–14896 (2008).
68. P. Zhang, *et al.*, The physical microenvironment of hematopoietic stem cells and its emerging roles in engineering applications. *Stem Cell Res. Ther.* **10**, 327 (2019).
69. F. Chowdhury, *et al.*, Material properties of the cell dictate stress-induced spreading and differentiation in embryonic stem cells. *Nat. Mater.* **9**, 82–88 (2010).
70. C. K. Materese, A. Savelyev, G. A. Papoian, Counterion atmosphere and hydration patterns near a nucleosome core particle. *J. Am. Chem. Soc.* (2009) https:/doi.org/10.1021/ja905376q.
71. A. T. Fenley, D. A. Adams, A. V. Onufriev, Charge state of the globular histone core controls stability of the nucleosome. *Biophys. J.* **99**, 1577–1585 (2010).
72. D. A. Potoyan, G. A. Papoian, Energy Landscape Analyses of Disordered Histone Tails Reveal Special Organization of Their Conformational Dynamics. *J. Am. Chem. Soc.* **133**, 7405–7415 (2011).
73. D. Winogradoff, I. Echeverria, D. A. Potoyan, G. A. Papoian, The acetylation landscape of the H4 histone tail: Disentangling the interplay between the specific and cumulative effects. *J. Am. Chem. Soc.* **137**, 6245–6253 (2015).
74. A. Davtyan, *et al.*, AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* (2012) https:/doi.org/10.1021/jp212541y.
75. B. Zhang, W. Zheng, G. A. Papoian, P. G. Wolynes, Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* **138**, 8126–8133 (2016).
76. H. Wu, Y. Dalal, G. A. Papoian, Binding Dynamics of Disordered Linker Histone H1 with a Nucleosomal Particle. *J. Mol. Biol.* **433**, 166881 (2021).
77. Y. Dalal, T. Furuyama, D. Vermaak, S. Henikoff, Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl. Acad. Sci.* **104**, 15974 LP – 15981 (2007).
78. M. Bui, *et al.*, Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell* **150**, 317–326 (2012).
79. M. P. Walkiewicz, E. K. Dimitriadis, Y. Dalal, CENP-A octamers do not confer a reduction in nucleosome height by AFM. *Nat. Struct. Mol. Biol.* **21**, 2–3 (2014).
80. D. Winogradoff, H. Zhao, Y. Dalal, G. A. Papoian, Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome. *Sci. Rep.* (2015) https:/doi.org/10.1038/srep17038.
81. J. Nye, D. Sturgill, R. Athwal, Y. Dalal, HJURP antagonizes CENP-A mislocalization driven by the H3.3 chaperones HIRA and DAXX. *PLoS One* **13** (2018).
82. S. Kmiecik, *et al.*, Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936 (2016).
83. E. F. Koslover, C. Fuller, A. F. Straight, A. J. Spakowitz, Role of DNA

Elasticity and Nucleosome Geometry in Hierarchical Packaging of Chromatin. *Biophys. J.* **98**, 474a (2010).
84. S. P. Hergeth, R. Schneider, The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* **16**, 1439–1453 (2015).
85. S. Sanulli, *et al.*, HP1 reshapes nucleosome core to promote heterochromatin phase separation. *Nature* (2019) https:/doi.org/10.1038/s41586-019-1669-2.
86. A. R. Shim, *et al.*, Dynamic Crowding Regulates Transcription. *Biophys. J.* **118**, 2117–2129 (2020).
87. A. Agrawal, N. Ganai, S. Sengupta, G. I. Menon, Nonequilibrium Biophysical Processes Influence the Large-Scale Architecture of the Cell Nucleus. *Biophys. J.* **118**, 2229–2244 (2020).
88. H. D. Ou, *et al.*, ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science (80-. ).* **357** (2017).
89. S. Shinkai, T. Sugawara, H. Miura, I. Hiratani, S. Onami, Microrheology for Hi-C Data Reveals the Spectrum of the Dynamic 3D Genome Organization. *Biophys. J.* **118**, 2220–2228 (2020).
90. E. H. Finn, *et al.*, Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e10 (2019).
91. Y. Takei, *et al.*, Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
92. J. Rodriguez, *et al.*, Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. *Cell* **176**, 213-226.e18 (2019).
93. T. Nagano, *et al.*, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
94. J. D. Buenrostro, *et al.*, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
95. A. Boettiger, S. Murphy, Advances in Chromatin Imaging at Kilobase-Scale Resolution. *Trends Genet.* **36**, 273–287 (2020).
96. B. R. Cairns, *et al.*, RSC, an Essential, Abundant Chromatin-Remodeling Complex. *Cell* **87**, 1249–1260 (1996).
97. Y. Lorch, B. Maier-Davis, R. D. Kornberg, Mechanism of chromatin remodeling. *Proc. Natl. Acad. Sci.* **107**, 3458 LP – 3462 (2010).
98. J. A. Armstrong, B. M. Emerson, Transcription of chromatin: these are complex times. *Curr. Opin. Genet. Dev.* **8**, 165–172 (1998).
99. M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12168–12173 (2016).
100. S. T. Spagnol, T. J. Armiger, K. N. Dahl, Mechanobiology of Chromatin and the Nuclear Interior. *Cell. Mol. Bioeng.* (2016) https:/doi.org/10.1007/s12195-016-0444-9.
101. B. Burke, Chain reaction: LINC complexes and nuclear positioning. *F1000Research* (2019) https:/doi.org/10.12688/f1000research.16877.1.
102. S. M. Schreiner, P. K. Koo, Y. Zhao, S. G. J. Mochrie, M. C. King, The tethering of chromatin to the nuclear envelope supports nuclear mechanics. *Nat. Commun.* (2015) https:/doi.org/10.1038/ncomms8159.
103. O. J. Rando, Combinatorial complexity in chromatin structure and function:

Revisiting the histone code. *Curr. Opin. Genet. Dev.* (2012) https:/doi.org/10.1016/j.gde.2012.02.013.
104. J. R. Davie, E. P. M. Candido, Acetylated histone H4 is preferentially associated with template active chromatin. *Proc. Natl. Acad. Sci. U. S. A.* (1978) https:/doi.org/10.1073/pnas.75.8.3574.
105. P. Ferrari, M. Strubin, Uncoupling histone turnover from transcription-associated histone H3 modifications. *Nucleic Acids Res.* (2015) https:/doi.org/10.1093/nar/gkv282.
106. J. Xie, *et al.*, Histone H3 Threonine Phosphorylation Regulates Asymmetric Histone Inheritance in the Drosophila Male Germline. *Cell* (2015) https:/doi.org/10.1016/j.cell.2015.10.002.
107. P. Tropberger, *et al.*, Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* (2013) https:/doi.org/10.1016/j.cell.2013.01.032.
108. F. Pryde, *et al.*, H3 K36 methylation helps determine the timing of Cdc45 association with replication origins. *PLoS One* (2009) https:/doi.org/10.1371/journal.pone.0005882.
109. N. L. Maas, K. M. Miller, L. G. DeFazio, D. P. Toczyski, Cell Cycle and Checkpoint Regulation of Histone H3 K56 Acetylation by Hst3 and Hst4. *Mol. Cell* (2006) https:/doi.org/10.1016/j.molcel.2006.06.006.
110. M. Manohar, *et al.*, Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *J. Biol. Chem.* (2009) https:/doi.org/10.1074/jbc.M109.003202.
111. J. Schneider, P. Bajwa, F. C. Johnson, S. R. Bhaumik, A. Shilatifard, Rtt109 is required for proper H3K56 acetylation: A chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* (2006) https:/doi.org/10.1074/jbc.C600265200.
112. S. J. Hainer, J. A. Martens, Identification of Histone Mutants That Are Defective for Transcription-Coupled Nucleosome Occupancy. *Mol. Cell. Biol.* (2011) https:/doi.org/10.1128/mcb.05195-11.
113. B. N. Devaiah, *et al.*, BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat. Struct. Mol. Biol.* (2016) https:/doi.org/10.1038/nsmb.3228.
114. T. Urahama, *et al.*, Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. *Epigenetics and Chromatin* (2016) https:/doi.org/10.1186/s13072-016-0051-y.
115. S. Giri, *et al.*, The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. *Elife* (2015) https:/doi.org/10.7554/eLife.06496.
116. D. Molina-Serrano, A. Kirmizis, Beyond the histone tail: Acetylation at the nucleosome dyad commands transcription. *Nucl. (United States)* (2013) https:/doi.org/10.4161/nucl.26051.
117. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* (2007) https:/doi.org/10.1038/nsmb1338.

118. A. Lagana, *et al.*, A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. *Nat. Cell Biol.* (2010) https:/doi.org/10.1038/ncb2129.
119. Y. Fujita, *et al.*, Priming of Centromere for CENP-A Recruitment by Human hMis18α, hMis18β, and M18BP1. *Dev. Cell* (2007) https:/doi.org/10.1016/j.devcel.2006.11.002.
120. A. O. Bailey, *et al.*, Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. *Proc. Natl. Acad. Sci. U. S. A.* (2013) https:/doi.org/10.1073/pnas.1300325110.
121. Y. Niikura, R. Kitagawa, K. Kitagawa, CENP-A Ubiquitylation Is Inherited through Dimerization between Cell Divisions. *Cell Rep.* (2016) https:/doi.org/10.1016/j.celrep.2016.03.010.
122. Y. Niikura, R. Kitagawa, K. Kitagawa, CENP-A Ubiquitylation Is Required for CENP-A Deposition at the Centromere. *Dev. Cell* (2017) https:/doi.org/10.1016/j.devcel.2016.12.020.
123. Z. Yu, *et al.*, Dynamic phosphorylation of CENP-A at Ser68 orchestrates its cell-cycle-dependent deposition at centromeres. *Dev. Cell* (2015) https:/doi.org/10.1016/j.devcel.2014.11.030.
124. C. W. Carroll, K. J. Milks, A. F. Straight, Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* (2010) https:/doi.org/10.1083/jcb.201001013.
125. S. Ramachandran, S. Henikoff, Replicating nucleosomes. *Sci. Adv.* (2015) https:/doi.org/10.1126/sciadv.1500587.
126. H. Kato, *et al.*, A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science (80-. ).* **340**, 1110–1113 (2013).
127. T. Kato, *et al.*, Activation of Holliday junction-recognizing protein involved in the chromosomal stability and immortality of cancer cells. *Cancer Res.* (2007) https:/doi.org/10.1158/0008-5472.CAN-07-1307.
128. B. H. Zimm, Theory of ``Melting'' of the Helical Form in Double Chains of the DNA Type. *J. Chem. Phys.* **33**, 1349–1356 (1960).
129. D. Shechter, H. L. Dormann, C. D. Allis, S. B. Hake, Extraction, purification and analysis of histones. *Nat. Protoc.* (2007) https:/doi.org/10.1038/nprot.2007.202.
130. S. B. Hake, *et al.*, Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem.* (2006) https:/doi.org/10.1074/jbc.M509266200.
131. J. H. Waterborg, A. J. Robertson, D. L. Tatar, C. M. Borza, J. R. Davie, Histones of Chlamydomonas reinhardtii: Synthesis, acetylation, and methylation. *Plant Physiol.* (1995) https:/doi.org/10.1104/pp.109.2.393.
132. A. G. Nuccio, M. Bui, Y. Dalal, A. Nita-Lazar, "Mass Spectrometry-Based Methodology for Identification of Native Histone Variant Modifications From Mammalian Tissues and Solid Tumors" in *Methods in Enzymology*, (2017) https:/doi.org/10.1016/bs.mie.2016.09.035.
133. J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein

binding within nucleosomes. *J. Mol. Biol.* (2011) https:/doi.org/10.1016/j.jmb.2011.01.003.
134. M. Brehove, *et al.*, Histone core phosphorylation regulates DNA accessibility. *J. Biol. Chem.* (2015) https:/doi.org/10.1074/jbc.M115.661363.
135. M. Bui, *et al.*, Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell* (2012) https:/doi.org/10.1016/j.cell.2012.05.035.
136. A. T. Fenley, R. Anandakrishnan, Y. H. Kidane, A. V. Onufriev, Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core. *Epigenetics and Chromatin* **11** (2018).
137. T. Panchenko, *et al.*, Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proc. Natl. Acad. Sci. U. S. A.* (2011) https:/doi.org/10.1073/pnas.1113621108.
138. C. P. Geiss, *et al.*, CENP-A arrays are more condensed than canonical arrays at low ionic strength. *Biophys. J.* (2014) https:/doi.org/10.1016/j.bpj.2014.01.005.
139. H. Fu, *et al.*, Methylation of Histone H3 on Lysine 79 Associates with a Group of Replication Origins and Helps Limit DNA Replication Once per Cell Cycle. *PLoS Genet.* (2013) https:/doi.org/10.1371/journal.pgen.1003542.
140. R. R. Yu, *et al.*, Hybrid phase ligation for efficient synthesis of histone proteins. *Org. Biomol. Chem.* (2016) https:/doi.org/10.1039/c5ob02195b.
141. M. J. Abraham, *et al.*, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* (2015) https:/doi.org/10.1016/j.softx.2015.06.001.
142. R. B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 (2009).
143. K. Lindorff-Larsen, *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
144. A. Pérez, *et al.*, Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **92**, 3817–3829 (2007).
145. I. S. Joung, T. E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
146. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* (1983) https:/doi.org/10.1063/1.445869.
147. a Šali, MODELLER: A Program for Protein Structure Modeling Release 9.12, r9480. *Rockefeller Univ.*, 779–815 (2013).
148. I. N. Shindyalov, P. E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747 (1998).
149. L. Schrödinger, The PyMol Molecular Graphics System, Versión 1.8. *Thomas Hold.* (2015).
150. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity

rescaling. *J. Chem. Phys.* **126**, 14101 (2007).
151. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* (1984) https:/doi.org/10.1063/1.448118.
152. G. Bussi, T. Zykova-Timan, M. Parrinello, Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **130** (2009).
153. K. Zhou, G. Gaullier, K. Luger, Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol.* (2019) https:/doi.org/10.1038/s41594-018-0166-x.
154. C. L. Peterson, G. Almouzni, Nucleosome dynamics as modular systems that integrate DNA damage and repair. *Cold Spring Harb. Perspect. Biol.* (2013) https:/doi.org/10.1101/cshperspect.a012658.
155. S. Web, *et al.*, The Centromere Paradox : Stable Inheritance with Rapidly Evolving DNA. *Science (80-. ).* **1098** (2009).
156. M. Di Pierro, D. A. Potoyan, P. G. Wolynes, J. N. Onuchic, Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7753–7758 (2018).
157. G. Hamilton, Y. Dimitrova, T. N. Davis, Seeing is believing: our evolving view of kinetochore structure, composition, and assembly. *Curr. Opin. Cell Biol.* (2019) https:/doi.org/10.1016/j.ceb.2019.03.016.
158. F. G. Westhorpe, A. F. Straight, The centromere: Epigenetic control of chromosome segregation during mitosis. *Cold Spring Harb. Perspect. Biol.* (2015) https:/doi.org/10.1101/cshperspect.a015818.
159. S. Müller, G. Almouzni, Chromatin dynamics during the cell cycle at centromeres. *Nat. Rev. Genet.* (2017) https:/doi.org/10.1038/nrg.2016.157.
160. P. B. Talbert, S. Henikoff, Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. *Trends Genet.* (2018) https:/doi.org/10.1016/j.tig.2018.05.001.
161. B. Stephanidis, S. Adichtchev, P. Gouet, A. McPherson, A. Mermet, Elastic Properties of Viruses. *Biophys. J.* **93**, 1354–1359 (2007).
162. A. P. Perrino, R. Garcia, How soft is a single protein? the stress-strain curve of antibody pentamers with 5 pN and 50 pm resolutions. *Nanoscale* (2016) https:/doi.org/10.1039/c5nr07957h.
163. B. Hess, Convergence of sampling in protein simulations. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* (2002) https:/doi.org/10.1103/PhysRevE.65.031910.
164. M. A. Balsera, W. Wriggers, Y. Oono, K. Schulten, Principal component analysis and long time protein dynamics. *J. Phys. Chem.* (1996) https:/doi.org/10.1021/jp9536920.
165. L. Y. Guo, *et al.*, Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat. Commun.* (2017) https:/doi.org/10.1038/ncomms15775.
166. S. J. Falk, *et al.*, CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science (80-. ).* (2015) https:/doi.org/10.1126/science.1259308.
167. S. J. Falk, *et al.*, CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. *Nat. Struct. Mol. Biol.*

(2016) https:/doi.org/10.1038/nsmb.3175.
168. A. E. García, Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* (1992) https:/doi.org/10.1103/PhysRevLett.68.2696.
169. A. Korostelev, H. F. Noller, Analysis of Structural Dynamics in the Ribosome by TLS Crystallographic Refinement. *J. Mol. Biol.* (2007) https:/doi.org/10.1016/j.jmb.2007.08.054.
170. M. Radmacher, M. Fritz, J. P. Cleveland, D. A. Walters, P. K. Hansma, Imaging Adhesion Forces and Elasticity of Lysozyme Adsorbed on Mica with the Atomic Force Microscope. *Langmuir* **10**, 3809–3814 (1994).
171. A. Vinckier, G. Semenza, Measuring elasticity of biological materials by atomic force microscopy in *FEBS Letters*, (1998) https:/doi.org/10.1016/S0014-5793(98)00592-4.
172. A. Parra, E. Casero, E. Lorenzo, F. Pariente, L. Vázquez, Nanomechanical properties of globular proteins: Lactate oxidase. *Langmuir* (2007) https:/doi.org/10.1021/la062864p.
173. T. Rakshit, S. Banerjee, S. Mishra, R. Mukhopadhyay, Nanoscale mechano-electronic behavior of a metalloprotein as a variable of metal content. *Langmuir* **29**, 12511–12519 (2013).
174. W. H. Roos, AFM nanoindentation of protein shells, expanding the approach beyond viruses. *Semin. Cell Dev. Biol.* (2018) https:/doi.org/10.1016/j.semcdb.2017.07.044.
175. W. F. Heinz, J. H. Hoh, Spatially resolved force spectroscopy of biological surfaces using the atomic force microscope. *Trends Biotechnol.* (1999) https:/doi.org/10.1016/S0167-7799(99)01304-9.
176. H. J. Butt, B. Cappella, M. Kappl, Force measurements with the atomic force microscope: Technique, interpretation and applications. *Surf. Sci. Rep.* (2005) https:/doi.org/10.1016/j.surfrep.2005.08.003.
177. D. P. Melters, T. Rakshit, S. A. Grigoryev, D. Sturgill, Y. Dalal, CENP-C stabilizes the conformation of CENP-A nucleosomes within the inner kinetochore at human centromeres. *bioRxiv*, 604223 (2019).
178. S. Kilic, *et al.*, Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nat. Commun.* (2018) https:/doi.org/10.1038/s41467-017-02619-5.
179. S. Sanulli, *et al.*, Disorganization of the histone core promotes organization of heterochromatin into phase-separated droplets. *bioRxiv* (2018) https:/doi.org/10.1101/473132.
180. K. Klare, *et al.*, CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J. Cell Biol.* (2015) https:/doi.org/10.1083/jcb.201412028.
181. K. J. Milks, B. Moree, A. F. Straight, Dissection of CENP-C-directed centromere and kinetochore assembly. *Proteins*.
182. M. R. Przewloka, *et al.*, CENP-C is a structural platform for kinetochore assembly. *Curr. Biol.* **21**, 399–405 (2011).
183. S. A. Ribeiro, *et al.*, A super-resolution map of the vertebrate kinetochore. *Proc. Natl. Acad. Sci.* **107**, 10484 LP – 10489 (2010).
184. B. D. Ross, *et al.*, Stepwise evolution of essential centromere function in a

Drosophila neogene. *Science (80-. ).* (2013) https:/doi.org/10.1126/science.1234393.
185. I. A. Drinnenberg, S. Henikoff, H. S. Malik, Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus? *Trends Cell Biol.* (2016) https:/doi.org/10.1016/j.tcb.2016.01.005.
186. J. J. van Hooff, E. Tromer, L. M. van Wijk, B. Snel, G. J. Kops, Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep.* (2017) https:/doi.org/10.15252/embr.201744102.
187. M. Abraham, *et al.*, User Manual. *SpringerReference* (2011) https:/doi.org/10.1007/springerreference_28001.
188. K. Lindorff-Larsen, *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* (2010) https:/doi.org/10.1002/prot.22711.
189. A. S. Lemak, N. K. Balabaev, On the berendsen thermostat. *Mol. Simul.* **13**, 177–187 (1994).
190. J. F. Marko, Micromechanical studies of mitotic chromosomes. *Chromosom. Res.* **16**, 469–497 (2008).
191. R. B. Nicklas, A quantitative study of chromosomal elasticity and its influence on chromosome movement. *Chromosoma* **14**, 276–295 (1963).
192. A. H. B. De Vries, B. E. Krenn, R. Van Driel, V. Subramaniam, J. S. Kanger, Direct observation of nanomechanical properties of chromatin in living cells. *Nano Lett.* **7**, 1424–1427 (2007).
193. C. Claudet, D. Angelov, P. Bouvet, S. Dimitrov, J. Bednar, Histone Octamer Instability under Single Molecule Experiment Conditions. *J. Biol. Chem.* **280**, 19958–19965 (2005).
194. D. C. Bouck, K. Bloom, Pericentric Chromatin Is an Elastic Component of the Mitotic Spindle. *Curr. Biol.* **17**, 741–748 (2007).
195. R. Collepardo-Guevara, T. Schlick, Insights into chromatin fibre structure by in vitro and in silico single-molecule stretching experiments. *Biochem. Soc. Trans.* **41**, 494–500 (2013).
196. D. A. Beard, T. Schlick, Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure* **9**, 105–114 (2001).
197. G. Shi, L. Liu, C. Hyeon, D. Thirumalai, Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun.* **9** (2018).
198. A. Tajik, *et al.*, Transcription upregulation via force-induced direct stretching of chromatin. *Nat. Mater.* **15**, 1287–1296 (2016).
199. P. Heun, *et al.*, Mislocalization of the drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev. Cell* **10**, 303–315 (2006).
200. J. Gosline, *et al.*, Elastic proteins: Biological roles and mechanical properties. *Philos. Trans. R. Soc. B Biol. Sci.* **357**, 121–132 (2002).
201. W. S. Klug, W. H. Roos, G. J. L. Wuite, Unlocking internal prestress from protein nanoshells. *Phys. Rev. Lett.* **109** (2012).
202. M. M. Gibbons, W. S. Klug, Influence of nonuniform geometry on nanoindentation of viral capsids. *Biophys. J.* **95**, 3640–3649 (2008).
203. J. T. Lloyd, J. A. Zimmerman, R. E. Jones, X. W. Zhou, D. L. McDowell,

Finite element analysis of an atomistically derived cohesive model for brittle fracture. *Model. Simul. Mater. Sci. Eng.* **19** (2011).
204. M. Zink, H. Grubmüller, Mechanical properties of the icosahedral shell of southern bean mosaic virus: A molecular dynamics study. *Biophys. J.* **96**, 1350–1363 (2009).
205. M. Zink, H. Grubmüller, Primary changes of the mechanical properties of southern bean mosaic virus upon calcium removal. *Biophys. J.* **98**, 687–695 (2010).
206. O. Kononova, K. A. Marx, V. Barsegov, " Nanoindentation in silico of Biological Particles " in *Applied Nanoindentation in Advanced Materials*, (2017), pp. 393–428.
207. M. Parrinello, A. Rahman, Strain fluctuations and elastic constants. *J. Chem. Phys.* **76**, 2662–2666 (1982).
208. F. Montalenti, *et al.*, Fully coherent growth of Ge on free-standing Si(001) nanomesas. *Phys. Rev. B - Condens. Matter Mater. Phys.* **89** (2014).
209. M. Salvalaglio, A. Voigt, K. R. Elder, Closing the gap between atomic-scale lattice deformations and continuum elasticity. *npj Comput. Mater.* **5** (2019).
210. M. Pitman, Minimal Cylinder Analysis (2020) https:/doi.org/10.5281/zenodo.2641192.
211. M. Abraham, *et al.*, GROMACS User Manual version 5.0.4. *www.Gromacs.org* (2014) https:/doi.org/10.1007/SpringerReference_28001.
212. M. K. Lee, K. H. Lee, S. H. Yoo, C. Y. Park, Impact of initial active engagement in self-monitoring with a telemonitoring device on glycemic control among patients with type 2 diabetes. *Sci. Rep.* **7** (2017).
213. O. Seksek, J. Bolard, Nuclear pH gradient in mammalian cells revealed by laser microspectrofluorimetry. *J. Cell Sci.* **109**, 257–262 (1996).
214. C. Chree, Longitudinal vibrations in solid and hollow cylinders. *Proc. Phys. Soc. London* **16**, 304–322 (1897).
215. L. R. W., The Theory of Sound. *Nature* **58**, 121–122 (1898).
216. K. Piaras, Solid Mechanics Part 1: An introduction to Solid Mechanics. *Solid Mech. Part I An Introd. to Solid Mech. Lect. notes*, 183–191 (2015).
217. P. D. T. O'connor, *Advanced strength and applied elasticity, A.C. Ugural and S.K. Fenster, Second SI Edition, Elsevier Science Publishing Company Inc., 1987. Number of pages: 471. Price: $35.00 (hardback only)* (1988).
218. A. Ahadi, D. Johansson, A. Evilevitch, Modeling and simulation of the mechanical response from nanoindentation test of DNA-filled viral capsids. *J. Biol. Phys.* **39**, 183–199 (2013).
219. M. T. Jaeger, Basic Relationships in Elasticity Theory. **53** Available at: https://dspace.mit.edu/bitstream/handle/1721.1/39137/22-314JSpring2004/NR/rdonlyres/Nuclear-Engineering/22-314JSpring2004/ACF104D9-8C23-4EEB-B3CD-D350871D8EDE/0/elastic_theory.pdf.
220. J. G. Williams, C. Gamonpilas, Using the simple compression test to determine Young's modulus, Poisson's ratio and the Coulomb friction coefficient. *Int. J. Solids Struct.* **45**, 4448–4459 (2008).
221. H. Hertz, Ueber die Berührung fester elastischer Körper. *J. fur die Reine und*

*Angew. Math.* **1882**, 156–171 (1882).
222. H. Xiao, *et al.*, Molecular basis of CENP-C association with the CENP-A nucleosome at yeast centromeres. *Genes Dev.* **31**, 1958–1972 (2017).
223. B. R. Zhou, *et al.*, Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol. Cell* **59**, 628–638 (2015).
224. T. Tian, *et al.*, Molecular basis for CENP-N recognition of CENP-A nucleosome on the human kinetochore. *Cell Res.* **28**, 374–378 (2018).
225. Y. Roulland, *et al.*, The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol. Cell* **63**, 674–685 (2016).
226. A. Hamiche, P. Schultz, V. Ramakrishnan, P. Oudet, A. Prunell, Linker histone-dependent DNA structure in linear mononucleosomes. *J. Mol. Biol.* **257**, 30–42 (1996).
227. J. Zlatanova, C. Seebart, M. Tomschik, The linker-protein network: control of nucleosomal DNA accessibility. *Trends Biochem. Sci.* **33**, 247–253 (2008).
228. R. T. Simpson, Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones. *Biochemistry* **17**, 5524–5531 (1978).
229. F. Cui, V. B. Zhurkin, Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Res.* **37**, 2818–2829 (2009).
230. J. Bednar, *et al.*, Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol. Cell* **66**, 384-397.e8 (2017).
231. N. Conde e Silva, *et al.*, CENP-A-containing Nucleosomes: Easier Disassembly versus Exclusive Centromeric Localization. *J. Mol. Biol.* **370**, 555–573 (2007).
232. S. L. McGovern, Y. Qi, L. Pusztai, W. F. Symmans, T. A. Buchholz, Centromere protein-A, an essential centromere protein, is a prognostic marker for relapse in estrogen receptor-positive breast cancer. *Breast Cancer Res.* **14** (2012).
233. T. Tomonaga, *et al.*, Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res.* **63**, 3511–3516 (2003).
234. Y. Mito, J. G. Henikoff, S. Henikoff, Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**, 1090–1097 (2005).
235. F. Mattiroli, *et al.*, DNA-mediated association of two histone-bound complexes of yeast chromatin assembly factor-1 (CAF-1) drives tetrasome assembly in the wake of DNA replication. *Elife* **6** (2017).
236. W. H. Liu, *et al.*, The Cac1 subunit of histone chaperone CAF-1 organizes CAF-1-H3/H4 architecture and tetramerizes histones. *Elife* **5** (2016).
237. S. H. Kim, R. Vlijm, J. Van Der Torre, Y. Dalal, C. Dekker, CENP-A and H3 nucleosomes display a similar stability to force-mediated disassembly. *PLoS One* **11** (2016).
238. O. Moreno-Moreno, M. Torras-Llort, F. Azorin, The E3-ligases SCFPpa and APC/CCdh1 co-operate to regulate CENP-ACID expression across the cell cycle. *Nucleic Acids Res.* **47**, 3395–3406 (2019).
239. D. J. Amor, *et al.*, Human centromere repositioning “in progress.” *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6542–6547 (2004).

240. P. Tessarz, T. Kouzarides, Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15**, 703–708 (2014).
241. K. Schwab, *The fourth industrial revolution* (World Economic Forum, 2016).
242. H. Gao, *et al.*, Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).
243. M. Basu, K. Wang, E. Ruppin, S. Hannenhalli, Predicting tissue-specific gene expression from whole blood transcriptome. *Sci. Adv.* **7**, eabd6991 (2021).
244. F. Aguet, *et al.*, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
245. S. Pressé, K. Ghosh, J. Lee, K. A. Dill, Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).
246. A. P. Latham, B. Zhang, Improving Coarse-Grained Protein Force Fields with Small-Angle X-ray Scattering Data. *J. Phys. Chem. B* **123**, 1026–1034 (2019).
247. X. Zhao, G. Plata, P. D. Dixit, SiGMoiD: A super-statistical generative model for binary data. *bioRxiv*, 2020.10.14.338277 (2021).
248. F. Morcos, *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci.* **108**, E1293 LP-E1301 (2011).
249. A. W. Senior, *et al.*, Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
250. H. Ashoor, *et al.*, Graph embedding and unsupervised learning predict genomic sub-compartments from HiC chromatin interaction data. *Nat. Commun.* **11**, 1173 (2020).
251. J. Berry, S. C. Weber, N. Vaidya, M. Haataja, C. P. Brangwynne, RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci.* **112**, E5237 LP-E5245 (2015).
252. K. Baumann, Human cohesin extrudes interphase DNA to make loops. *Nat. Rev. Mol. Cell Biol.* **21**, 3 (2020).
253. Q. MacPherson, B. Beltran, A. J. Spakowitz, Chromatin Compaction Leads to a Preference for Peripheral Heterochromatin. *Biophys. J.* **118**, 1479–1488 (2020).
254. B. Burke, C. L. Stewart, The nuclear lamins: flexibility in function. *Nat. Rev. Mol. Cell Biol.* **14**, 13–24 (2013).
255. M. Pitman, D. P. Melters, Y. Dalal, Job Opening for Nucleosome Mechanic: Flexibility Required. *Cells* (2020) https:/doi.org/10.3390/cells9030580.
256. M. P. Walkiewicz, M. Bui, D. Quénet, Y. Dalal, Tracking histone variant nucleosomes across the human cell cycle using biophysical, biochemical, and cytological analyses. *Methods Mol. Biol.* (2014) https:/doi.org/10.1007/978-1-4939-0888-2_34.
257. E. K. Dimitriadis, C. Weber, R. K. Gill, S. Diekmann, Y. Dalal, Tetrameric organization of vertebrate centromeric nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20317–20322 (2010).
258. M. P. Stumme-Diers, S. Banerjee, M. Hashemi, Z. Sun, Y. L. Lyubchenko, Nanoscale dynamics of centromere nucleosomes and the critical roles of CENP-A. *Nucleic Acids Res.* **46**, 94–103 (2018).
259. J. L. Hutter, J. Bechhoefer, Calibration of atomic-force microscope tips. *Rev.*

*Sci. Instrum.* **64**, 1868–1873 (1993).